Assessing the Quality of Quality Assessment: The Role of Scheduling

HARVARD | BUSINESS | SCHOOL

Assessing the Quality of Quality Assessment: The Role of Scheduling

Maria R. Ibanez
Michael W. Toffel

Working Paper 17-090

Assessing the Quality of Quality Assessment: The Role of Scheduling

Maria R. Ibanez
Harvard Business School

Michael W. Toffel
Harvard Business School

Working Paper 17-090

Assessing the Quality of Quality Assessment: The Role of Scheduling†

Maria R. Ibanez
Harvard Business School, Boston, MA 02163, mibanez@hbs.edu,

Michael W. Toffel
Harvard Business School, Boston, MA 02163, mtoffel@hbs.edu

April 14, 2017

Many production processes are subject to inspection to ensure they meet quality, safety, and environmental standards imposed by companies and regulators. This paper explores how the scheduling of inspections risks introducing bias that erodes inspection quality by altering inspector stringency. We theorize that inspection results will be affected by (a) when the inspection occurs within an inspector's daily schedule and (b) the inspection outcomes of the inspector's prior inspected establishment. Analyzing thousands of food safety inspections, we find that inspectors cite fewer violations in successive inspections throughout their day and when inspections risk prolonging their typical workday. We also find that inspectors cite more violations after inspecting establishments that exhibited worse compliance or greater compliance deterioration. We discuss several implications for managers who schedule or rely on inspections.

Keywords: quality; assessment; bias; inspection; scheduling; econometric analysis; empirical research; regulation

1. Introduction

Many companies inspect their own and their suppliers' operations to ensure they are meeting quality, labor, and environmental standards. Various government agencies also conduct inspections for regulatory compliance. The accuracy of inspections is critical to their being a useful input to key managerial decisions, including how to allocate quality improvement resources, which suppliers to source from, and how to penalize compliance failures. Inaccurate assessments can prevent managers, workers, customers, and neighbors from making well-informed decisions based on the risks imposed by an establishment's operations. Moreover, inspections that miss what they could have caught can undermine the inspection regime's ability to deter intentional noncompliance. In this study, we theorize and find evidence of several sources of bias that lead to inaccurate inspections. We also propose solutions—including

† We gratefully acknowledge helpful comments and support from numerous health officers and coordinators, Xiang Ao, Ryan Buell, Matthew Grochowski, Matthew Johnson, Andrew Marder, Melissa Ouellet, Ananth Raman, Brad Staats, and participants in the Harvard Business School TOM Workshop, and financial support from Harvard Business School's Division of Research and Faculty Development.

alternative inspection scheduling regimes—that can improve inspection accuracy without increasing inspection costs.

Several studies have revealed various sources of inspection inaccuracy. As most of these studies focus on general errors, little is yet known about inspector bias. We consider bias that results from an operational decision: scheduling, an unexplored type of bias. Building on work from the behavioral sciences, we hypothesize how the sequence of inspections might affect the number of violations cited. Specifically, inspector stringency on a particular inspection may be influenced by (1) its position within the day (daily schedule effects) and (2) the outcomes of the inspector's prior inspection (prior inspection outcome effects, or, simply, outcome-effects). (Throughout this paper, we refer to *an inspector's* preceding inspection as his or her "prior" inspection, and *an establishment's* preceding inspection as its "previous" inspection.)

We study the influence of scheduling on inspection accuracy in the context of local health department food safety inspections of restaurants and other food-handling establishments. While these inspections need to accurately assess compliance in order to protect consumer health, the number of violations cited in these reports is a function of both the facility's actual hygiene at the time of the inspection and the inspector's stringency in detecting and recording violations. Using data on thousands of inspections, we find strong evidence that inspectors' schedules affect the violations cited in inspection reports.

We hypothesize two daily schedule effects. We first theorize that each additional inspection over the course of a day causes fatigue that erodes inspectors' stringency, which leads them to cite fewer violations. We find empirical evidence to support this, observing that each subsequent inspection during an inspector's day yields 3.15% fewer violation citations, an effect that our supplemental analysis demonstrates is not due to inspectors' scheduling presumably cleaner establishments later in their workday. Second, we hypothesize that inspections that risk prolonging an inspector's workday will be conducted less stringently, which will lead to fewer violations being cited. We find empirical support for this, in that potentially shift-prolonging inspections yield 5.07% fewer violation citations.

We then hypothesize three ways in which an inspector's experience at one inspection affects the number of violations cited at his or her next inspection. First, we hypothesize that an inspector's stringency will be influenced by the number of violations at his or her prior inspection. Those violations will affect the inspector's emotions and perceptions about the general compliance of the community of inspected establishments (via the salience of those recent inspection results), in turn altering his or her expectations and attitudes when inspecting the next establishment. This leads us to predict that having just conducted an inspection that cites more violations will lead the inspector to also cite more violations in the next establishment he or she inspects. As predicted, we find that *each additional violation* cited in the inspector's prior inspection (of a different establishment) increases by 1.51% the number of violations that inspector cites at the next establishment he or she inspects.

We hypothesize that trends matter, too: discovering more compliance deterioration (or less improvement) at one inspection affects inspectors' emotions and perceptions in ways that lead them to cite more violations at the next establishment they visit. Supporting this hypothesis, we find that inspectors cite 1.31% more (fewer) violations after having inspected another establishment whose violation trend worsened (improved) by one standard deviation. Finally, we hypothesize that, based on negativity bias, this trend effect will be stronger following an inspection that found deterioration than one that found improvement. Indeed, we find empirical evidence that the trend effect is asymmetric, applying when compliance at the inspector's prior establishment deteriorates but not when it improves.

Our work contributes to both theory and practice. By identifying factors that bias inspections, our work contributes to the literature on monitoring and quality improvement (e.g., Gray, Siemsen, and Vasudeva 2015). Our focus on how scheduling affects inspector stringency introduces the operational lens of scheduling to the literature examining inspector bias, which has otherwise largely focused on experience or other sociological and economic factors (e.g., Short, Toffel, and Hugill 2016, Ball, Siemsen, and Shah 2017). Our examination of how operational decisions affect inspector behavior also contributes to the literature on behavioral operations, which emphasizes the importance of human behavior in operations management decisions (Bendoly, Donohue, and Schultz 2006, Gino and Pisano

2008). Our findings show that fatigue can affect performance of primary tasks even during normal shift hours. Moreover, by examining data from actual decisions with important consequences for public health, we contribute to the recent attempts to explore high-stakes decision-making in field settings (e.g., Chen et al., 2016). With managers across many different industries seeking to monitor and improve quality, our research suggests a cost-effective lever: exploiting the behavioral effects of the organization of work.

2. Related Literature

Our research builds on three streams of literature: monitoring organizations' adherence to operational standards, the impact of inspectors' efficacy and biases on inspection accuracy, and scheduling and task sequencing as drivers of task performance.

2.1. Monitoring and Assessment of Standards Adherence

Decades of scholarship have explored various approaches to ensuring that operations adhere to specifications provided by internal engineering and quality control departments, customers, and regulators. These approaches include statistical process control (e.g., Porteus and Angelus 1997), total quality management (e.g., Lapré, Mukherjee, and Van Wassenhove 2000), programs that encourage operators to self-disclose process errors and regulatory violations (Leape 2002, Gawande and Bohara 2005, Toffel and Short 2011, Kim 2015), and electronic monitoring systems such as radio frequency identification (RFID) (Staats et al. 2016). Physical inspection remains a primary approach; for example, internal quality control departments assess manufacturing processes (Shah, Ball, and Netessine 2016) and internal auditors assess inventory records (Kök and Shang 2007). Some companies hire third-party monitors to assess their suppliers' operations (Handley and Gray 2013, Locke 2013, Short and Toffel 2016), including whether they are complying with standards such as the ISO 9001 quality management system standard (Corbett 2006, Levine and Toffel 2010, Gray, Anand, and Roth 2015), in part to protect the buying companies' reputation. An extensive literature has highlighted the role of inspections in promoting operational routines and adherence to legally required Good Manufacturing Processes (e.g., Anand, Gray, and Siemsen 2011, Gray, Siemsen, and Vasudeva 2015), occupational health and safety regulations (e.g., Ko, Mendeloff, and Gray 2010, Levine, Toffel, and Johnson 2012), and environmental regulations (for a review, see Shimshack 2014). Inspections have also been shown to have durable and cumulative beneficial effects on process control and compliance, with successive inspections yielding continuous improvement in working conditions as establishments resolve the concerns identified in inspections (Ko, Mendeloff, and Gray 2010, Toffel, Short, and Ouellet 2015).

Given that inspections are costly, organizations use a variety of approaches to determine which establishments to prioritize. Several studies have explored tradeoffs between various targeting regimes. Some regimes seek to maximize the number of violations detected by prioritizing establishments suspected of being the worst performers (Harrington 1988, Kang et al. 2013), others prioritize establishments that have gone the longest without being inspected, and still others target all establishments with equal probability based on randomization (e.g., Lana 2003, Johnson, Levine, and Toffel 2017). Whereas those studies can help managers understand which establishments to prioritize over, say, the course of a year, little is known about the relative effectiveness of different approaches to sequencing inspections, the topic of our study.

2.2. Inspection Accuracy

The usefulness of inspections is contingent on their accuracy. Researchers have long been interested in how to conduct quality control inspections (e.g., Ballou and Pazer 1982), recognizing inspectors' fallibility and variability (Feinstein 1989). The limited number of studies of the heterogeneity across inspectors' propensity to report violations has identified the importance of their tenure, training, gender, and former exposure to the establishment (Macher, Mayo, and Nickerson 2011, Short, Toffel, and Hugill 2016, Ball, Siemsen, and Shah 2017). Inspector scrutiny among third-party inspection firms has been shown to be influenced by (a) whether it is the inspected establishment or another party that hires the inspection firm and pays for the inspection (Ronen 2010, Duflo et al. 2013, Short and Toffel 2016), (b) the level of competition among inspection firms (Bennett et al. 2013), and (c) whether the inspecting firm has cross-selling opportunities (Koh, Rajgopal, and Srinivasan 2013, Pierce and Toffel 2013). In contrast to these demographic aspects of individual inspectors and structural dimensions of the relationship between the inspection firm and the inspected establishment, we explore a very different potential source of inspection bias: where the inspection falls within an inspector's schedule.

2.3. The Drivers of Task Performance

The literature examining task performance has focused mostly on how workers vary their pace in response to workload. Heavier workloads prompt workers to speed up (reducing service times) (Schultz et al. 1998, Kc and Terwiesch 2009, Kc and Terwiesch 2012, Tan and Netessine 2014, Delasay, Ingolfsson, and Kolfal 2016) and to shift tasks upstream (Batt and Terwiesch 2016) or downstream (Freeman, Savva, and Scholtes 2017). Extended periods of excessive workload can cause workers to slow down (Kc and Terwiesch 2009) and can erode service quality (Oliva and Sterman 2001). Heavier workloads can also lead workers to conduct fewer tasks (Oliva and Sterman 2001, Kc and Terwiesch 2012, Kuntz, Mennicken, and Scholtes 2015) and to conduct them less comprehensively, which can manifest as lower service quality (Kuntz, Mennicken, and Scholtes 2015, Berry Jaeker and Tucker 2016) and as incomplete documentation resulting in lost revenues (Powell, Savin, and Savva 2012). In contrast to workload, we focus on work schedule—in particular, examining how task sequence, tasks scheduled near the end of shift, and prior tasks affect service quality in the form of conducting comprehensive inspections.

Research on the role of scheduling on task performance has investigated the optimal allocation of labor to tasks over time and has addressed problems such as machine inspection scheduling (Lee and Rosenblatt 1987) and workforce scheduling (Green, Savin, and Savva 2013). Studies of task sequencing have shown, for example, that scheduling similar tasks consecutively can improve performance by increasing task repetition and reducing delays incurred from switching tasks (e.g., Staats and Gino 2012, Ibanez et al. 2016). Recent work has begun to explore how task sequencing within a shift affects worker behavior and performance. For example, two recent studies find that healthcare workers work more quickly later in the shift (Deo and Jain 2015, Ibanez et al. 2016). In contrast to their focus on task speed, we focus on task quality in a setting that purports to provide consistent quality inspections as the basis for a fair and objective monitoring regime.

Another study finds that healthcare workers become less compliant with handwashing rules over the course of their shift (Dai et al. 2015). That study focused on adherence to a secondary task that was largely unobservable to others, where noncompliance was common, and where fatigue might lead workers to shift their attention from the secondary task toward their primary tasks. In contrast, our study focuses on primary tasks and on the outcome of such tasks—the number of violations cited—that is explicitly observable to others. Such visibility could deter variation. Moreover, whereas Dai et al. (2015) measured adherence dichotomously, we use a more nuanced scalar measure.

A few recent articles based on psychology and behavioral economics have examined how various decisions are affected by the number of prior decisions in a given day and the outcomes of those decisions. One study of eight judges found that (a) they were more likely to deny parole (that is, to preserve the status quo) as they issued more judgments throughout the course of their day, suggesting that repeated decisions caused mental depletion, and (b) this bias was attenuated by a food break (Danziger, Levav, and Avnaim-Pesso 2011). A study of MBA application assessments found that the higher the cumulative average of the scores interviewers had given to applicants at a given moment on a given day, the lower the interviewers scored subsequent applicants that day, suggesting that decision makers exhibit bias to maintain a consistent daily acceptance rate (Simonsohn and Gino 2013). A third study found that judges, loan reviewers, and baseball umpires were more likely to make "accept" decisions immediately after a "reject" decision (and vice versa), exhibiting negative autocorrelation that results in decision errors (Chen, Moskowitz, and Shue 2016).[1]

---

[1] These studies discuss several mechanisms. Danziger, Levav, and Avnaim-Pesso (2011) argue that that the act of making decisions leads to mental depletion, which in turn increases the likelihood of easier decisions—rejecting requests. Simonsohn and Gino (2013) argue that their evidence is consistent with narrow bracketing; that is, decision makers dividing decisions into daily subsets and avoiding deviations from an expected overall acceptance rate for each subset. The authors propose three possible mechanisms: mental accounting (making daily decisions in line with a long-term evaluation target); the law of small numbers (underestimating the likelihood of sequential streaks occurring by chance); and pleasing supervisors believed to expect subordinates' decisions to follow this pattern. The authors rule out contrast effects; that is, prior cases serving as reference points in the evaluation of the current case in which lower evaluations follow higher-quality cases and higher evaluations follow lower-

Because inspectors are trained to be accurate, it is unclear whether they will be vulnerable to time or sequence effects. Moreover, if they are affected, the effects might be the opposite of what prior studies have shown. First, whereas judges were shown to become harsher as they made more decisions throughout the day, it is not obvious whether inspector behavior will follow suit, given that inspector harshness (which manifests as stringency) requires additional effort. Second, though prior work finds a negative correlation of decisions with the running average of prior decisions and with the latest decisions, inspectors do not have explicit or self-imposed quotas or targets and, as we explain below, their emotions and perceptions may be affected by their prior tasks in ways that promote positive correlation over time. Additionally, we go beyond what prior work has considered by proposing that the magnitude of the effects from prior task outcomes will be asymmetric, depending on whether those outcomes were positive or negative.

3. Theory and Hypotheses

Quality assurance audits and inspections have detailed procedures to be followed in pursuit of accuracy. Yet, in practice, behavioral biases may influence an inspector's stringency. Whereas inspections are typically assumed to yield the same results no matter when they occur on the inspector's schedule, we hypothesize that inspection results will indeed be influenced by when an inspection occurs during an inspector's daily schedule—which we refer to as *daily schedule effects*—and by the type of experience inspectors have at their immediately prior inspection—which we refer to as *prior inspection outcome effects*.

3.1. Daily Schedule Effects on Quality Assessment

3.1.1. Inspector Fatigue. Inspectors' work typically consists of a sequence of evaluative tasks that include physical tasks (such as manually examining the dimensions of a part or the temperature of a freezer) and mental tasks (such as interviewing an employee or determining whether a set of observations

---

quality cases. Chen, Moskowitz, and Shue (2016) conclude that the errors are consistent with the law of small numbers (or gambler's fallacy) or sequential contrast effects and less consistent with quotas, learning, or fairness preferences.

is or is not within acceptable standards). As these tasks are executed, physical and mental fatigue will increase (Brachet, David, and Drechsler 2012). Furthermore, experimental evidence indicates that mental fatigue increases physical fatigue (Wright et al. 2007, Marcora, Staiano, and Manning 2009).

Over the course of a day, inspectors' physical and mental fatigue will reduce their physical and cognitive effort. This undermines stringency, which requires physical and cognitive efforts such as moving throughout the facility, interviewing personnel, waiting to observe work, executing procedures such as taking measurements, and conducting unpleasant tasks in undesirable conditions (such as observing storage practices in a walk-in freezer). Once an attribute is observed, inspectors need to recall and interpret the relevant standards to decide whether there is a violation and, if so, to document it. Each step must be executed according to rules that increase the complexity even of tasks that might appear simple to the untrained eye. Moreover, mental effort is required to make decisions against the status quo; as inspectors grow more tired during the day, they may become more willing to accept the status quo (Muraven and Baumeister 2000, Danziger, Levav, and Avnaim-Pesso 2011), which in the context of inspections can take the form of passing inspection items. Finally, mental effort is also required to withstand the social confrontations that can erupt when a finding of noncompliance is disputed by those working at the establishment, who may genuinely disagree and for whom, in any case, much may be at stake in terms of reputation and sales. Citing violations can also provoke threats of appeals and lawsuits. Anticipating such responses, as they grow more fatigued, inspectors will exert less effort and seek to avoid confrontation, both of which contribute to their exhibiting more leniency. For all these reasons, we hypothesize:

Hypothesis 1: Inspectors will cite fewer violations as they complete more inspections throughout their daily schedule.

3.1.2. Potential Shift Prolonging. Inspectors' behavior may also be affected by whether, as they approach the end of the shift, they begin an inspection that would be anticipated to last beyond when they typically end work for the day. Because inspections require visiting the establishment, it is inefficient for the inspector to suspend an inspection once underway; the inspector would then have to bear the travel cost again the next day to finish. The inability to interrupt an inspection, combined with a desire to finish at their typical time, may create pressure to speed up and inspect less thoroughly. In many settings, workers have discretion over their pace, which can lead them to prolong tasks to fill the time available (Hasija, Pinker, and Shumsky 2009) or else to conduct work more quickly when they perceive their workload to be higher than usual (Kc and Terwiesch 2012; Berry-Jaeker and Tucker 2016) and to avoid working unpaid hours beyond their shift (Chan 2015). As workers approach their typical end-of-shift time, accomplishing their remaining workload can become increasingly pressing as their perceived opportunity cost of time increases. Their desire to speed up work in these circumstances can result in the increased reliance on workarounds and cutting corners (Oliva and Sterman 2001) which, in turn, can reduce the quality of the work performed. Because properly conducted inspections require carefully evaluating a series of individual elements to identify whether each is in or out of compliance, omitting or expediting tasks to avoid prolonging their shift will result in a less comprehensive inspection and, consequently, fewer violations will be detected and cited. We therefore hypothesize:

Hypothesis 2: Inspectors will cite fewer violations at inspections when they are at risk of working beyond their end of shift.

3.2. Prior Inspection Outcome Effects on Quality Assessment

3.2.1. Violation Levels of the Inspector's Prior Inspection. Inspectors are influenced not only by the sequencing of inspections within the day, but also by the results of prior inspections. One type of these outcome-effects is driven by whether the establishment an inspector just visited had many or few violations. There are two reasons why inspecting an establishment with many violations can imbue inspectors with a negative attitude that leads them to inspect more diligently at their next inspection, whereas inspecting a more compliant establishment can lead them to be less stringent in their subsequent inspection.

First, an inspector's prior inspection can affect him or her emotionally. Personnel at that prior establishment are likely to be more (less) dissatisfied and resentful when more (fewer) violations are cited and inspectors may absorb those emotions through emotional contagion (Barsade 2002), as well as through interactions that are more (less) hostile (Neuman and Baron 1997). This, in turn, can influence inspectors' goodwill and thus their stringency during the next inspection.

Second, the experience at the inspector's prior inspection can shape his or her perceptions regarding the overall behavior of establishments, which can influence his or her stringency at the subsequent inspection. Recently experiencing an event (such as compliance) increases its salience and results in more rapid recall. An inspector may therefore use the results of that inspection to update his or her estimate of typical compliance levels, relying on the availability heuristic (Tversky and Kahneman 1974) and seeking evidence at his or her next inspected establishment that supports these expectations, consistent with confirmation bias (Nickerson 1998). This becomes a self-fulfilling prophecy, as the inspector's heightened or reduced scrutiny detects more or fewer violations. We therefore hypothesize:

Hypothesis 3: The more (fewer) violations an inspector cites at one establishment, the more (fewer) violations he or she will cite in the next establishment.

3.2.2. Violation Trend at the Inspector's Prior Inspection. An inspector's behavior is shaped not only by the prior establishment's *level* of compliance, but also by its *change* in compliance relative to its previous inspection. This second type of outcome-effect also results from how the prior inspection affects the inspector's emotions and perceptions.

The inspector's emotional response (through emotional contagion and interactions) at his or her prior establishment will depend on the trend there because the expectations of the establishment's personnel will be based on its previous inspection; they will be pleased or displeased according to whether its violation count has decreased or increased. After visiting an establishment with greater improvement, the inspector will exhibit a more positive temperament and will approach his or her next inspection with greater empathy and less stringency.

The inspectors' perceptions, too, may be biased by the change in violations at the inspector's prior establishment. Many inspectors view inspections as a cooperative endeavor with the regulated entity to help improve business operations and safeguards stakeholders (e.g., May and Wood 2003, Pautz 2009, Pautz 2010). Improved compliance may therefore be attributed to management taking the rules and regulations seriously—that is, cooperating—whereas worsened compliance may be attributed to management ignoring or deliberately flouting the rules—definitely not cooperating. Improved compliance therefore confirms a cooperative relationship, which can lead inspectors to believe that the overall community of inspected establishments is cooperating and thus become less stringent in the next inspection. Worsened compliance can lead inspectors to believe that the overall community of inspected establishments is not cooperating and thus become more stringent in the next inspection. We therefore hypothesize:

> Hypothesis 4: The more an establishment's compliance has deteriorated (improved), the more (fewer) violations an inspector will record at the next establishment.

3.2.3. Violation Trend at the Inspector's Prior Inspection: Asymmetric Effects of Deterioration versus Improvement. According to the principle of *negativity bias*, negative events are generally more salient and dominant than positive events (Rozin and Royzman 2001). Negative events instigate greater information processing to search for meaning and justification, which in turn strengthens the memory and tends to spur stronger and more enduring effects in many psychological dimensions (Baumeister et al. 2001).

Negativity bias can affect the impact of the prior inspection's violation trend on the inspector's emotions and perceptions. First, negativity bias implies that for the inspected establishment's staff, the negative emotional effect of a drop in compliance may be stronger than the positive emotional effect of an improvement. This would result in a stronger conveyance to inspectors of negative emotions associated with a drop in compliance and a weaker conveyance of positive emotions associated with an improvement. An inspector will then absorb more negative emotions after the negative finding than positive emotions after the positive finding. Moreover, as argued by Barsade (2002), mood contagion might be more likely for unpleasant emotions because of higher attention and automatic mimicry. These asymmetries in the extent to which declining versus improving conditions affect inspectors' emotions will lead, in turn, to asymmetric effects on the strength of the resulting positive versus negative outcome-effects.

Second, the salience of negative outcomes may have a stronger effect on inspectors' perceptions about how all of the establishments they monitor generally think about compliance, which can shape their stringency in a subsequent inspection. This is due to the *status-quo bias*: with the status quo acting as the reference point, negative changes are perceived as larger than positive changes of the same magnitude (Samuelson and Zeckhauser 1988, Kahneman 2003). We therefore hypothesize:

Hypothesis 5: Observing deteriorated conditions at an establishment will increase the inspector's stringency at the next establishment to a greater extent than observing improved conditions will reduce his or her stringency.

4. Empirical Analysis

4.1. Empirical Context: Food Safety Inspections

Our hypotheses are ideally tested in an empirical context in which inspectors (a) conduct multiple inspections per day and (b) work individually, which avoids the challenge of discerning individuals' behaviors from those of co-inspectors. Food safety inspections—in which environmental health officers working for local health departments inspect restaurants and other food-handling establishments to protect consumers by monitoring compliance and educating kitchen managers—fulfill both criteria. Moreover, food safety inspections, commonly known as restaurant health inspections, are designed to minimize foodborne illness; noncompliance can jeopardize consumer health. The quality of these assessments—and their ability to safeguard public health—depends on the accuracy of inspectors.

Foodborne disease in the US is estimated to cause 48 million illnesses resulting in 128,000 hospitalizations and 3,000 deaths each year, imposing billions of dollars of medical costs and costs associated with reduced productivity and with pain and suffering (Scallan et al. 2011, Scharff 2012, Minor et al. 2015). Violations can affect firms' reputations and revenues and can trigger organizational responses that range from additional training for responsible personnel to legal representation to refute citations.

Several prior studies have examined food safety inspections. For example, Lehman, Kovács, and Carroll (2014) found that consumers are less concerned about food safety at restaurants they perceive to be more "authentic." Others have investigated the extent to which restaurants improve hygiene practices once they were required to disclose their inspection results to consumers via restaurant grade cards (Jin and Leslie 2003, Simon et al. 2005, Jin and Leslie 2009). More recent studies have found that online customer reviews of restaurants contain text that reflects hygiene conditions that can predict health inspections results (Kang et al. 2013) and can increase inspector effectiveness if health inspection agencies take them into account when prioritizing establishments for inspection (Glaeser et al. 2016).

Because inspectors need evidence to justify citing violations (and thus can only cite violations if they are truly present), studies of inspection bias (e.g., Bennett et al. 2013, Duflo et al. 2013, Short, Toffel, and Hugill 2016) are based on the assumption that deviations from the true number of violations are only due to underdetection, and that bias does not lead inspectors to cite violations that are not actually present. This assumption was validated in our interviews with inspectors, and underlies our empirical approach. Moreover, because violations are based on regulations regarding food operations based on science-based guidance for protecting consumers, each violation item is relevant.

We purchased data from Hazel Analytics, a company that gathers restaurant inspections from several local governments across the United States, processes the information to create electronic datasets, and sells these datasets to researchers and to companies—such as restaurant chains—interested in monitoring their licensees. These datasets include information about the inspected establishment (name, identification number, address, city, state, ZIP code), the inspector, the inspection type, the date, the times when the inspection began and ended, the violations recorded, and, where available, the inspector's comments on those violations.

We purchased all of Hazel Analytics' inspection datasets that included inspection start and end times as well as unique identifiers for each inspector, all of which are necessary to observe inspector schedules. This included all food safety inspections conducted in Lake County, Illinois from September 4, 2013, to October 5, 2015; Camden County, New Jersey from September 4, 2012, to September 24, 2015;

and in the state of Alaska from December 8, 2007, to October 4, 2015. (These date ranges reflect all inspections from these domains that Hazel Analytics had coded.) Our estimation sample omits (a) inspector-days for which we cannot adequately calculate relevant variables based on what appear to be data entry errors that we were unable to correct (for example, when there was ambiguity about inspection sequence) and (b) inspections that are dropped by our conditional fixed-effects Poisson specification. Details are provided in Appendix A. This results in an estimation sample containing 12,017 inspections of 3,399 establishments conducted by 86 inspectors on 6,880 inspector-days in Camden County, New Jersey (1,402 inspections), Lake County, Illinois (8,962 inspections), and the state of Alaska (1,653 inspections). (These sample restrictions do not affect our inferences, as all of our hypothesized results continue to hold when using alternative specifications estimated on all inspections in the raw dataset, as described in Appendix A.)

Our interviews with managers and inspectors at health inspection departments represented in our dataset indicate that inspectors have limited discretion over scheduling their inspections. Each inspector is responsible for inspecting all establishments within his or her assigned geographic territory[2] and inspectors are rotated to different territories every two or three years. Inspectors are instructed to schedule their inspections by prioritizing establishments based on their due dates, which are computed based on previous inspection dates and the required inspection frequency for each establishment type.[3] To minimize travel time, inspectors are instructed to group inspections with similar due dates by geographic proximity.

Though inspectors also carry out many administrative duties (such as reviewing records, answering emails, and attending department meetings at the office), the main components of their work are the inspections and the associated travel. Inspectors typically review the most recent inspections as

[2] Temporary exceptions are provided in special situations, such as staffing shortages, sick leave, or vacation.
[3] Required inspection frequency varies based on the riskiness of an establishment's operations. For example, those selling only prepackaged foods impose lower risk of foodborne illness and are thus required to be inspected less often than those that handle raw ingredients or that prepare and store hot or cold food more than 12 hours before serving. This prioritization is facilitated by software that health inspection departments use to track inspections, which allows sorting of the establishments to be inspected by due date.

they prepare for their next inspections. Traveling between their office and establishments to inspect represents a substantial portion of inspectors' days because of the geographical dispersion in the areas covered by our data. Inspectors are discouraged from working overtime.

4.2. Measures

4.2.1. Dependent and Independent Variables. We measure *violations* as the total number of violations cited in each inspection, a typical approach used by others (e.g., Helland 1998, Stafford 2003, Langpap and Shimshack 2010, Short, Toffel, and Hugill 2016).

We measure an inspector's schedule-induced fatigue at a given inspection as the *number of prior inspections today*, computed as the number of inspections that the inspector had already conducted before the focal inspection on the same day. Thus, this variable is coded 0 for an inspector's first inspection of the day, 1 for the second, and so on.[4]

To measure whether an inspection might reasonably be anticipated to conclude after the inspector's typical end-of-shift time, we created an indicator variable, *potentially shift-prolonging*, coded 1 when the anticipated end time of an inspection (calculated as the inspection start time plus the duration of that establishment's previous inspection) falls after the inspector's running average daily clock-out time based on all of that inspector's preceding days in our sample, and coded 0 otherwise.

Prior inspected establishment's violations is the number of violations the inspector cited at the establishment inspected prior to the focal inspection, whether minutes or days earlier.

Prior inspected establishment's violation trend is calculated as the percentage change in the number of violations at that establishment between that day's inspection and its previous inspection (we added one to the denominator to avoid dividing by zero).

We create two indicator variables to distinguish whether the inspector's prior establishment had improved, deteriorated, or not substantially changed its number of violations compared to its previous

[4] We also conduct our analyses measuring schedule-induced fatigue in three alternative ways: the inspector's actual, anticipated, or predicted cumulative minutes already spent that day conducting inspections. See the Robustness Tests section for details.

inspection. We classify an establishment's violation trend as *improved saliently* (or *deteriorated saliently*) if its current inspection yielded at least two fewer (more) violations than its previous inspection. (The intermediate case, in which the number of violations differed by only one or remained constant, serves as the baseline condition.) We create the dummy variables *prior inspected establishment saliently improved*, coded 1 when the inspector's prior inspected establishment *improved saliently* and 0 otherwise, and *prior inspected establishment saliently deteriorated*, coded 1 when the inspector's prior inspected establishment *deteriorated saliently* and 0 otherwise. An inspection conducted immediately after the inspection of an establishment whose performance change was only one or no violations is considered the baseline condition.

4.2.2. Control Variables. We measure *inspector experience* as the cumulative number of inspections the inspector had conducted (at any establishment) since the beginning of our sample period by the time he or she began the focal inspection.

We create an indicator variable, *returning inspector*, coded 1 when the inspector of the focal inspection had inspected the establishment beforehand and 0 otherwise.

We create two indicator variables to designate the time of day the inspection began: *breakfast period* (midnight to 10:59 am) and *dinner period* (4:00 pm–11:59 pm), with the remaining *lunch period* (11:00 am–3:59 pm) serving as (omitted) baseline condition. We also create a series of indicator variables specifying the month and the year of the inspection.

We create a series of indicator variables to control for whether the inspection is the *establishment's nth inspection (second through tenth or more)*, each of which indicates whether an inspection is the establishment's first, second, third (and so on) inspection in our sample period.

We create a series of inspection-type dummies to indicate whether the inspection was routine, routine-education, related to permitting, due to a complaint, an illness investigation, or a follow-up. *Routine inspections* are conducted to periodically monitor establishments; *routine-education inspections* are particular cases of routine inspections in which an educational presentation is conducted to train establishment staff. These two types represent 79% of the inspections in our estimation sample. *Permit* *inspections* are conducted when establishments change ownership or undergo construction, upgrades, or remodeling. *Complaint inspections* are triggered by the local health department receiving a complaint; *complaint risk inspections* are those that might have been so triggered.[5] *Illness investigation inspections* are those conducted to investigate a possible foodborne illness (food poisoning). A *follow-up inspection* (or re-inspection) is conducted to verify that violations in a preceding inspection have been corrected and thus is of limited scope. *Other inspections* (the omitted category) include visits to confirm an establishment's deactivation/closure, temporary events such as outdoor festivals, mobile establishments, and vending machines; this serves as the omitted category in our empirical specifications.

Tables 1 and 2 provide summary statistics and reports correlations.[6]

4.3. Empirical Specification

We test our hypotheses by estimating the following model:

$$Y_{ijen} = F(\beta_1 \eta_{ij} + \beta_2 \lambda_{ij} + \beta_3 \rho_{i,j-1} + \beta_4 \delta_{i,j-1} + \beta_5 \varphi_j + \beta_6 \mu_{ijen} + \beta_7 \tau_{ijen} + \beta_8 \nu_n + \beta_9 \gamma_{ijen} + \beta_{10} IE_{ie} + \varepsilon_{ijen}),$$

where Y_{ijen} is the number of *violations* cited in the *n*th inspection of establishment *e* that was conducted by inspector *i* and that was his or her *j*th inspection in our sample. $F(\cdot)$ refers to the Poisson function.

η_{ij} is inspector *i*'s *number of prior inspections today*. λ_{ij} refers to whether the inspection was *potentially shift-prolonging.* $\rho_{i,j-1}$ is the inspector's *prior inspected establishment's violations*; that is, the number of violations that inspector *i* cited at the immediately preceding inspection of another establishment. $\delta_{i,j-1}$ refers to the *prior inspected establishment's violation trend* or, in some specifications, the two variables that indicate particular ranges of that variable: *prior inspected establishment saliently improved* and *prior inspected establishment saliently deteriorated*.

[5] The Camden data does not explicitly code as "complaint inspections" those that were conducted in response to a complaint. We were able to identify which Camden inspections were *at risk* of being triggered by complaints, and coded these as *complaint risk inspections*. Specifically, we obtained a list from the Camden County Department of Health and Human Services that includes the dates and times when the department received a complaint and the name of the inspector who was sent to investigate it, but not the identity of the establishment. These inspectors investigate these complaints either the day they are received or the next day. Because we do not know which specific inspection corresponds to a complaint, we categorize as *complaint risk inspection* all (so-called) routine inspections that the assigned inspector conducted later in the day when the complaint was received and for all inspections that the inspector conducted the next day. In the few instances in which Camden's complaint list did not identify the inspector assigned, we categorize as *complaint risk inspection* all Camden routine inspections conducted later in the day when those complaints were received and throughout the next day.

[6] Supplemental descriptive statistics are provided in Table B1 in Appendix B.

We include φ_j to control for *inspector experience* (Short, Toffel, and Hugill 2016). We control for *returning inspector* (μ_{ijen}) because inspectors who return to an establishment they had inspected before tend to behave differently than inspectors who are new to the establishment (Short, Toffel, and Hugill 2016, Ball, Siemsen, and Shah 2017).

The vector τ_{ijen} includes *breakfast period* and *dinner period* to control for the possibility that an establishment's cleanliness might vary over the course of a day and because prior research indicates that many individual behaviors are affected by time of day (Linder et al. 2014, Dai et al. 2015).[7] τ_{ijen} also includes fixed effects for the month of the inspection and fixed effects for the year of the inspection.[8]

We include a series of fixed effects to control for the *establishment's nth inspection (second through tenth or more)* (ν_n) because research has shown that other types of establishments improve compliance over subsequent inspections (Ko, Mendeloff, and Gray 2010, Toffel, Short, and Ouellet 2015).

To accommodate the possibility that different types of inspection might mechanically result in different numbers of violations (e.g., due to different scope), the model includes *inspection type* dummies (γ_{ijen}).

Finally, we include fixed effects for every inspector-establishment combination (IE_{ie}). These inspector-establishment dyads control for all time-invariant inspector characteristics (such as gender and formal education) and all time-invariant establishment characteristics (such as cuisine type and geographic domain). Thus, our specification identifies changes in the number of violations that a particular inspector cited when inspecting a given establishment on different occasions. Note that this approach is more conservative than including separate sets of fixed effects for inspectors and for

[7] Results are robust to using hourly dummies, except that the *potentially shift-prolonging* effect is no longer statistically significant, which is not surprising because it is related to the end-of-shift times.
[8] Results are robust to including a series of indicator variables for the day of the week.

establishments.[9] By including inspector-establishment fixed effects, we focus on variation within an establishment-inspector dyad, which avoids concerns that spatial correlation is driving our results.[10]

4.4. Identification

Empirical support for our hypothesis that inspector fatigue reduces stringency (H1) would manifest as inspections conducted later in an inspector's daily schedule exhibiting fewer violations. The results of some preliminary analyses bolster our claim that such a trend is driven by inspector fatigue rather than two alternative explanations. One potential alternative explanation is that daily trends in customer visits and staffing levels could result in temporal trends in staff cleaning effort that might result in establishments exhibiting better hygiene conditions as the day wears on. The inspectors we interviewed indicated that it would be unlikely for hygiene conditions to routinely improve throughout the day and that, in fact, hygiene conditions may often get worse throughout the day as more customers are served. They also noted that many violations reflect poor procedures that are unlikely to shift throughout the day—such as poor labeling of packages and poor documenting procedures—and improperly functioning sinks. We nonetheless control for potential variation in establishments' cleanliness at different periods of the day by including fixed effects for time of day.

A second potential alternative explanation is that inspectors might intentionally arrange their daily schedules to begin with the establishments they expect will yield more violations, leaving the easier establishments for later in the day. Even if establishments with more expected violations were routinely scheduled earlier in the day by some inspectors, our inclusion of establishment-inspector-dyad fixed effects controls for time-invariant characteristics of establishments and inspectors. Moreover, we conducted the following analyses that enable us to rule out that establishments were being sequenced in

---

[9] We find similar results when we estimate an alternative model including separate sets of fixed effects for inspectors and for establishments. Because estimating this two-way fixed effects model using Poisson regression led to convergence problems, we instead used ordinary least squares regression with a log-transformed dependent variable (specifically, log of violations after adding 1).

[10] Specifically, our approach addresses potential concerns that spatial correlation might be driving our hypothesized (H3) positive correlation in violations cited at establishments that inspectors visit sequentially. Proximate establishments that inspectors tend to visit sequentially might exhibit similar violation counts because they share neighborhood characteristics that might affect the supply of and demand for compliance.

this manner. In particular, our dataset enables us to observe establishments' previous inspections, which inspectors also use to assess historical inspection outcomes. We therefore looked for evidence of such a trend by examining whether an inspector tended to start his or her shift with establishments that had more violations in their previous inspection, as our interviews with inspectors indicated they anticipated such establishments being especially likely to have violations in subsequent inspections. We used two approaches to examine this. First, we conducted a test to assess whether an establishment's number of violations at its previous inspection was related to the sequence in which it was subsequently inspected. After tabulating the number of violations from an establishment's previous inspection against its focal inspection sequence, a Pearson's chi-squared test indicated that previous violations were not significantly related to the focal inspection sequence ($\chi^2 = 139$, $p = 1.00$). We also created a box-and-whisker plot to visually confirm this result (see Figure B1 in Appendix B). This plot reveals stable medians and interquartile ranges of establishments' previous violations irrespective of the establishments' focal sequence and thus provides no visual evidence that inspectors created their daily schedules based on establishments' previous violation counts. Second, we estimated a Poisson regression of inspection sequence on number of violations from that establishment's previous inspection and inspector-day fixed effects, clustering standard errors by inspector-day. Our results ($\beta = 0.009$, S.E. = 0.003, $p < 0.01$) indicate that this effect is trivial: establishments with one standard deviation more violations in their previous inspection (that is, 2.5 violations) were, in their subsequent inspection, scheduled 0.02 later in an inspector's sequence (calculated as the product of the variable's mean and coefficient, or 2.5 * 0.009 = 0.02). The inverse of this coefficient indicates that, on average, inspectors would schedule an establishment one step later in their daily sequence if it had 111 more violations (44 standard deviations) than another establishment—a number that is outside the rage of outcomes and thus impossible. This tiny effect enables us to rule out that inspectors, to any meaningful degree, intentionally sequence the day's inspections based on difficulty. Moreover, the direction of this small effect enables us to rule out that inspectors schedule their more difficult inspections earlier in their day. This indicates that, if anything, the ordering of inspections would bias *against* our hypothesized effect.

Collectively, the results of our investigating these two concerns (a) indicate that none of these factors are threats to our identification, and (b) increase the confidence with which we can identify the effects of inspector fatigue on stringency.

We next consider identification issues with respect to H2. Our hypothesis that an inspection being *potentially shift-prolonging* will reduce the number of violations cited raises a potential concern that inspectors might intentionally schedule as *potentially shift-prolonging* those inspections that they anticipate will yield fewer violations. We used two approaches to examine this. First, we calculated the number of violations at each establishment's previous inspection and found this average to be 3.1 for establishments whose subsequent inspection was *potentially shift-prolonging* and 2.3 for establishments whose subsequent inspection was not. A Pearson's chi-squared test indicates that the difference in these distributions is statistically significant ($\chi^2 = 243$, $p < 0.01$), which provides strong evidence that inspectors did not intentionally schedule establishments expected to have *fewer* violations as potentially shift-prolonging. Second, we estimated a conditional logistic regression of the *potentially shift-prolonging* indicator on the number of violations from that establishment's previous inspection and inspector-day fixed effects, clustering standard errors by inspector-day. The results ($\beta = 0.104$, S.E. = 0.013, $p < 0.01$, odds ratio = 1.11) indicate that an additional violation in the establishment's previous inspection slightly increases the odds of being *potentially shift-prolonging*, which would only be a bias against our hypothesized effect. Collectively, these results increase the confidence with which we can identify the hypothesized effect of an inspection being *potentially shift-prolonging* on the number of violations cited.

Finally, we consider the identification of the prior inspection outcome effects hypothesized in H3-H5. If inspectors intentionally scheduled their inspections based on their expected outcomes (for example, by inspecting those they expect to have high violations in one week and those they expect to have fewer violations in another week), such sorting could drive our findings. Our interviews with managers and inspectors suggest that this is not the case and that inspections are grouped by location to minimize travel time because travel represents a large proportion of an inspector's time and thus cost. While it is possible that violations are spatially correlated (due, for example, to community economic demographics), our inclusion of establishment-inspector-dyad fixed effects controls for such time-invariant characteristics of the establishments. In other words, our identification strategy is based on the focal establishment's *deviation* from its average violations cited by a particular inspector. Moreover, our inclusion of month and year fixed effects controls for contemporaneous shocks (such as economic cycles) that might affect establishments' willingness or ability to manage hygiene.

4.5. Results

4.5.1. Model Results. We estimate the count model using fixed-effects Poisson regression and report standard errors clustered by establishment (Table 3).[11] Our results are robust to several alternatives: clustering standard errors by inspector, estimating the model with negative binomial regression with conditional fixed effects, and estimating the model using ordinary least squares regression predicting log violations. Multicollinearity is not a serious concern, given that variance inflation factors (VIFs) are less than 1.7 for all hypothesized variables and less than 6.1 for all variables except three of the inspection-type indicators. Because our specifications control for a variety of factors that affect the number of violations cited, we interpret coefficients on the hypothesized variables as evidence of bias, the same approach used by prior studies (e.g., Chen, Moskowitz, and Shue 2016, Short, Toffel, and Hugill 2016). Because deviations from the true number of violations are assumed to result only from underdetection (as described above), we interpret negative coefficients to indicate the extent of underdetection occurring, whereas positive coefficients indicate the extent to which underdetection is avoided. We interpret effect sizes based on incidence rate ratios (IRRs).

We test Hypotheses 1-4 using Model 1. We begin by interpreting the coefficients on our control variables. The estimated coefficient on inspector experience is positive and statistically significant, suggesting that, all else constant, the number of violations cited per inspection increases as the inspector conduct inspections over time, albeit by a small amount on an inspection-by-inspection level. The

[11] Poisson panel estimators are consistent even if the data are not Poisson distributed, provided the conditional mean is correctly specified (Gourieroux, Monfort, and Trognon 1984, Azoulay, Graff Zivin, and Wang 2010, Cameron and Trivedi 2010). Because of the weaker distributional assumption of the Poisson panel estimators, they may be more robust than negative binomial regression (Cameron and Trivedi 2010).

negative and statistically significant coefficient on *returning inspector* ($\beta = -0.116$, $p < 0.01$) indicates that inspectors who return to an establishment cite 11% fewer violations than inspectors who had not inspected that establishment beforehand, which is also consistent with prior studies. Considering time-of-day effects, we note that, on average, inspections conducted earlier in the day cite 6% more violations than inspections conducted during the lunch period, whereas inspections conducted during the dinner and lunch periods cite statistically indistinguishable numbers of violations. The estimated coefficients on the establishment's nth inspection (not reported) indicate that fewer violations were cited at successive inspections of a given establishment, a result consistent with prior research that has examined other types of inspections. For example, the estimated coefficient on the dummy variable denoting an establishment's third inspection ($\beta = -0.209$, $p < 0.01$) indicates that those inspections cite 19% fewer violations on average than its initial inspection.

We find support for daily schedule effects. The negative, statistically significant coefficient on *number of prior inspections today* ($\beta = -0.032$, $p < 0.01$) indicates that inspections conducted later in an inspector's daily schedule yield fewer violations. This suggests that fatigue reduces stringency, providing support for H1. The coefficient magnitude indicates that each subsequent inspection during the day results on average in 3.15% fewer violations cited. The results also indicate that inspections that risked extending an inspector's workday yielded fewer violations, as predicted by H2. The negative statistically significant coefficient on *potentially shift-prolonging* ($\beta = -0.052$, $p < 0.05$) indicates that such inspections result in 5.07% fewer citations.

To explore the influence of the outcome of the inspector's prior inspected establishment, we first consider the number of violations cited in that inspection. The coefficient on *prior inspected establishment's violations* is positive and statistically significant ($\beta = 0.015$, $p < 0.01$), which supports H3. Each additional citation at the establishment inspected immediately before the focal inspection increases the number of violations cited in the focal inspection by 1.51%. The statistically significant positive coefficient on *prior inspected establishment's violation trend* ($\beta = 0.013$, $p < 0.05$) supports H4.

A one-standard-deviation increase in this trend increases the number of citations in the focal inspection by 1.31%. Note that this is in addition to the effect of the number of violations (H3).

To test H5, Model 2 replaces *prior inspected establishment's violation trend* with the indicator variables *prior inspected establishment saliently improved* and *prior inspected establishment saliently deteriorated*. The baseline condition occurs when the establishment the inspector visited immediately prior to the focal inspection exhibited no more than one violation more or less than it did in its previous inspection. Compared to this baseline condition, our results indicate that inspectors cite more violations in inspections conducted after their prior inspected establishment exhibits salient deterioration ($\beta = 0.075$, $p < 0.01$). The IRR indicates that, on average, an inspector who has just inspected an establishment that exhibited salient deterioration will report 8% more violations in the focal inspection. However, we find no evidence that observing salient improvement in the prior inspected establishment has any effect on the number of violations cited in the focal inspection. A Wald test indicates that these effects significantly differ (Wald $\chi^2 = 4.21$, $p < 0.05$), which supports H5: observing salient deterioration in the prior inspected establishment has a statistically significant stronger spillover effect than observing salient improvement does on an inspectors' next inspection.

4.5.2. Results Interpretation. To better understand the magnitude of the estimated effects, we estimate the impacts of daily schedule effects and outcome-effects on violations for the average inspection, scale up the results across the entire United States, and translate the results into healthcare outcomes and costs. We report these results in Table 4.

We first consider the average impact per inspection. Specifically, we compare the status quo (that is, the current practice with its associated scheduling effects) with six alternative scenarios described in each row of Table 4, each of which mitigates one or more of the effects we identified. We make all these comparisons based on Model 1 in Table 3. We calculate average predicted values under each scenario based on the model's estimates after recoding to zero, one at a time, *number of prior inspections today, potentially shift-prolonging*, *prior inspected establishment's violations*, and *prior inspected* *establishment's violation trend*, and then simultaneously recoding to zero the first two of these variables (related to daily schedule effects), and then the latter two (outcome-effects). These results are reported in Column 1 of Table 4. For the status quo, we use the model's estimates to calculate the average predicted number of violations per inspection, based on actual values of all variables, to be 2.42365. Column 2 reports the percent change in the average predicted violations that compares each of these scenarios to the status quo. This shows the average impact per inspection, which takes into account both the magnitude of the effect(s) and their frequency.

Consider the scenario depicted in row 1, in which we estimate the violations and hospitalizations that would occur in the absence of the fatigue affect associated with prior inspections conducted in a given day. We nullify this effect by recoding to zero *number of prior inspections today* for all observations (and preserve the actual values for all other variables). Using the estimates from Model 1 of Table 4, we then calculate the average predicted number of violations to be 2.498 (Column 1). As noted in Column 2, this scenario would result in 3.1% more violations cited per inspection than the status quo of 2.42365 violations cited per inspection (Column 2).[12] This same approach suggests that eliminating the *potentially shift-prolonging* effect would increase violations cited per inspection by 1.4%. Thus, even though the estimated coefficient magnitudes described in the prior section indicate that, on average, each subsequent inspection during the day results in 3.15% fewer violations being cited per inspection and that *potentially shift-prolonging* inspections result in 5.07% fewer violations being cited per inspection, we find that eliminating the *number of prior inspections today* effect would have a larger impact than eliminating the *potentially shift-prolonging* effect.

We then estimate the potential nationwide implications of our calculations based on the assumption that the estimated one million food establishments that are monitored by state, local, and tribal agencies in the United States[13] are each inspected annually, and that our sample of inspections is

[12] We calculate 3.1% as (2.49802 - 2.42365) / 2.42365.
[13] US Food and Drug Administration, "Retail Food Protection," http://www.fda.gov/food/guidanceregulation/retailfoodprotection/, last updated Sep. 9, 2016, accessed Jan. 1, 2017.

typical of those conducted across the country. To calculate a nationwide figure, we take the difference between the average predicted values from each scenario and the status quo (that is, the Column 1 figure minus 2.42365), and multiply that by the one million inspections conducted annually across the country (Column 3).

Returning to the first scenario, which considers the impact of eliminating the *number of prior inspections today* effect, we scale the difference in average predicted violations per inspection that arise in this scenario compared to the status quo (2.49802 - 2.42365) by the one million inspections conducted annually nationwide, which yields 74,370 violations nationwide per year. Thus, if the inspector fatigue associated with prior inspections in a given workday could be eliminated, we estimate that 74,370 additional violations per year nationwide that are currently undetected would be cited (Column 3, row 1). This same approach leads us to estimate that eliminating the *potentially shift-prolonging* effect we identified would result in the citation of 33,020 undetected violations per year nationwide (row 2). Turning to the outcome effects, this approach indicates that 94,110 violations per year nationwide would not have been cited in the absence of the effect of *prior inspected establishment's violations*, and 18,750 would not have been cited in the absence of the effect of *prior inspected establishment's violation trend* (rows 4 and 5).

To evaluate the implications of these estimated nationwide violation counts associated with schedule effects, we translate these results into health outcomes, as well as associated costs. First, we convert the effects on violations compared to the status quo into a percent change in foodborne illness hospitalizations (Column 4). We do so by multiplying the percent change in the average predicted number of violations between the scenario and the status quo (Column 2) by the ratio of 20% decrease in foodborne illness hospitalizations per 5% improvement in restaurant compliance scores based on prior research of Los Angeles restaurants (Jin and Leslie 2003). For example, to assess the overall effect of inspector's fatigue from prior work earlier in the day (row 1), we multiply the 3.1% increase in the number of violations cited per inspection described earlier (Column 2) by the ratio of 20% decrease in hospitalizations per 5% improved restaurant score, which indicates that hospitalizations would decrease by 12.3% if inspector fatigue associated with prior work in a workday were eliminated. Similarly, hospitalizations would have been 5.5% lower in the absence of the potentially shift-prolonging effect (calculated as 1.4% * [-20% / 5%]); row 3).

We then estimate the difference in annual nationwide foodborne illness hospitalizations (Column 5) and annual nationwide foodborne illness cases (Column 6) that would arise in the absence of each effect, as well as the cost implications (Column 7). Our calculations are based on (1) the estimated annual 128,000 hospitalizations and 47.8 million foodborne illness cases that actually occurred (Scallan et al. 2011) and (2) two alternative estimates of the average cost per foodborne illness of $747 (Minor et al. 2015) and $1,626 (Scharff 2012) to construct the lower and upper bound of our total cost estimates. To estimate the impact on foodborne illness hospitalizations (Column 5), we compare the estimated number of nationwide annual hospitalizations that would have occurred in each scenario to the 128,000 hospitalizations that actually occurred, and calculate the difference by multiplying the rate of change (in Column 4) by the 128,000 hospitalizations. Because only 128,000 of the 47.8 million annual foodborne illnesses result in hospitalization, we compute the impact on illnesses (Column 6) by multiplying the estimated impact on hospitalizations (Column 5) by the ratio of 47.8 million to 128,000 (that is, 373.4 foodborne illness cases per foodborne illness hospitalization). For example, eliminating the effects of the *number of prior inspections today* (row 1) would result in 15,711 fewer hospitalizations, a 12.3% decline from the status quo of 128,000 hospitalizations, and 5.87 million fewer foodborne illness cases (row 1). Multiplying this 5.87 million fewer foodborne illness cases that would result (Column 6) by the $747 or $1,626 cost per case, we estimate that $4.383 to $9.540 billion in foodborne illness costs could be saved per year if the effects of the *number of prior inspections today* were eliminated nationwide.

We similarly estimate that the combined effect of the two daily schedule effects we identified (number of prior inspections today and potentially shift-prolonging effects; see row 3) is resulting in 109,140 undetected violations per year nationwide, which are associated with 23,056 excess foodborne illness hospitalizations, 8.61 million excess foodborne illness cases, and $6,432 to $14,000 million in estimated costs. Such unreported violations and excess illnesses and costs could be avoided if steps were taken to mitigate inspector fatigue, such as by restricting inspectors to a single inspection per day. Any such policies to mitigate fatigue should be pilot tested and the benefits weighed against the potential additional labor costs they might impose. We also estimate that 110,800 violations per year nationwide are being detected thanks to the combined effect of the two outcome effects we identified (prior inspected establishment's violations and prior inspected establishment's violation trend), which are preventing 23,407 foodborne illness hospitalizations and 8.74 illness cases, with an estimated cost of $6,529 to $14,213 million (row 6).

4.6. Robustness Tests

We conduct several analyses to confirm the robustness of our findings. Our primary results are based on a conservative approach that includes establishment-inspector-dyad fixed effects. We find similar results whether we instead include establishment fixed effects or separate sets of fixed effects for inspectors and for establishments (see footnote 9 for details). Also, to assess whether unusually busy days, which might lead inspectors to become especially fatigued, might be driving our schedule-induced fatigue (H1) results, we reestimated our models on the subsample of inspector-days with no more than six inspections (the 99th percentile). Our hypothesized results are robust to these subsample tests.

Our results regarding the effects of schedule-induced fatigue (H1) hold even when we measure this construct using any of the following three alternative approaches rather than the number of prior inspections on the day of the focal inspection. In our first alternative, we calculate the *actual cumulative minutes* inspectors spent onsite in their prior inspections that day to better account for the fact that inspection duration can differ and that longer (and not just more numerous) inspections are likely to cause more fatigue. Our second alternative approach accommodates the potential concern that fatigue increased the duration of prior inspections. Here, we calculate the *anticipated cumulative minutes* inspectors would expect to have spent onsite in their prior inspections that day, computed as the average of the durations of those establishments' previous two inspections (or their single previous inspection if only one is available). In our third alternative approach, we compute the *predicted cumulative minutes* inspectors would spend onsite in their prior inspections that day, using the predicted durations derived from an ordinary least squares regression model, with a log transformed outcome variable and including the covariates from the corresponding main specification.

Our results are also robust to including as additional controls in our primary models indicator variables denoting the day of the week the inspection occurred. Our results are mostly robust to substituting our three time-of-day periods (*breakfast period, lunch period,* and *dinner period*) with indicator variables for each hour of the day at which the inspection occurred. Only the *potentially shift-prolonging* coefficient is no longer statistically significant, likely due to the higher multicollinearity introduced by this approach.

4.7. Extensions

In this section, we extend our primary analysis along several dimensions. We first test whether outcome-effects persist beyond the inspector's very next inspection and beyond the same day. We then investigate the extent to which these effects influence the citing of different types of violation, distinguishing between critical and noncritical violations. We also consider how other aspects of inspections are affected, focusing on inspection duration, citation pace, and comments documenting violations.

4.7.1. Persistence of Outcome-effects. Our primary models examine the extent to which an inspector's experience at the prior establishment influences the number of violations cited at his or her next inspection. To explore whether these outcome-effects persist longer, we added to our models the *penultimate inspected establishment's violations* (that is, two establishments ago) and then also the *antepenultimate inspected establishment's violations* (that is, three establishments ago). Appendix Table C2 reports the results in Columns 1 and 2, respectively, which indicate that the number of violations cited in an inspection significantly affects the next three inspections; the declining coefficient magnitudes suggest that the effect dissipates.

We also assess whether outcome-effects persist as strongly after inspectors have the opportunity to take an overnight break. To compare between-day versus within-day spillovers, we split the *prior* *inspected establishment's violations* into the following two variables. *Prior inspected establishment's violations for the first inspection of the day* equals *prior inspected establishment's violations* for the first inspection of the day and equals 0 for all inspections conducted later in the day. *Prior inspected establishment's violations for the second+ inspection of the day* is coded oppositely: it equals *prior inspected establishment's violations* for all inspections other than the first of the day and equals 0 for the first inspection of the day. Estimating our primary model but substituting these two variables for *prior inspected establishment's violations* (Column 3 of Appendix Table C2) yields coefficients on *prior inspected establishment's violations for the first inspection of the day* ($\beta = 0.016$, $p < 0.01$) and *prior inspected establishment's violations for the second+ inspection of the day* ($\beta = 0.015$, $p < 0.01$) that are statistically indistinguishable (Wald $\chi^2 = 0.03$, $p = 0.86$). This indicates that outcome-effects persist to the same extent whether or not successive inspections occur on the same day or across different days.

4.7.2. Critical versus Noncritical Violations. To assess whether our hypothesized relationships differentially influence inspectors' behavior across different types of violations, we estimated our models on two subsets of violations. First, we predict the number of *critical violations*, which are related to food preparation practices and employee behaviors that more directly contribute to foodborne illness or injury. These factors are prioritized in Alaska and in Camden County by being displayed on the first page of the inspection report and in Lake County by being tagged in the reports. Second, we estimated our models on the number of *noncritical violations* (that is, violations of procedures often referred to as "good retail practices"). While less risky than the other type, these are also important for public health and include overall sanitation and preventative measures to protect foods, such as proper use of gloves. Inspections averaged 0.93 *critical violations* and 1.49 *noncritical violations* (see Table C1 for other summary statistics of these variables).

Appendix Table C3 reports the results. The estimated coefficients on the controls highlight the heterogeneity across these two types of violation. Though, on average, more noncritical violations are cited in inspections conducted during the breakfast period, there is no evidence that time of day affects critical violations. This is consistent with critical violations being related to longer-term establishment practices which are insensitive to the number of customers being served or the staff's ability to respond to the inspector's presence.

We find that the daily-schedule effects identified in our primary results are driven by noncritical violations rather than critical violations. In particular, we find no evidence that citations of critical violations are affected by daily-schedule affects: the coefficients on *number of prior inspections today* and *potentially shift-prolonging* are not statistically significant (Columns 1 and 2). This suggests that fatigue does not affect inspectors' ability to discover and report critical violations. In contrast, an inspector's daily schedule has large statistically significant effects on noncritical violations (Columns 3 and 4). The negative, statistically significant coefficients on *number of prior inspections today* (β = -0.041, $p < 0.01$) indicate that each subsequent inspection during the day results, on average, in 4.02% fewer noncritical violations being cited. The negative statistically significant coefficient on *potentially shift-prolonging* (β = -0.060, $p < 0.05$) indicates that such inspections result in 5.82% fewer citations.

Outcome-spillovers are more ubiquitous, affecting critical and noncritical violations alike. The magnitudes of the positive, statistically significant coefficients on *prior inspected establishment's violations* indicate that, on average, each additional violation cited at the inspector's prior inspected establishment is associated with 1.82% more critical violations (Column 1: β = 0.018, $p < 0.01$) and 1.41% more noncritical violations (Column 3: β = 0.014, $p < 0.01$) cited in the focal inspection.

As with total violations, there is no evidence of critical and noncritical violations being affected by the inspection being conducted after salient improvement was observed in the inspector's prior inspected establishment, that is, when the *prior inspected establishment saliently improved*. When the *prior inspected establishment saliently deteriorated*, inspections yield, on average, 7.36% more critical violations (Column 2: β = 0.071, $p < 0.10$) and 7.79% more noncritical violations (Column 4: β = 0.075, $p < 0.05$).

Overall, these results indicate that inspectors' schedules have somewhat different effects on citing critical versus noncritical violations. Citing noncritical violations appears to be influenced by both daily schedule effects and outcome-spillovers, while citing critical violations appears to be influenced only by outcome-spillovers.

4.7.3. Inspector Speed and Citation Pace. Our primary results show how daily schedules and inspections of prior establishments are associated with the number of violations cited. To assess whether such results might be driven by inspectors spending less time and exhibiting less scrutiny in the subsequent (focal) inspection, we estimate our primary models on the log of *inspection duration*, the number of minutes between an inspection's start time and end time. Moreover, to assess the net of these two effects, we explore the inspector's citation pace—a measure of productivity in this setting—and estimate our primary models on the log (after adding 1) of *violation citations per hour*. The results are reported in Table C4.

We find that inspectors conduct inspections more quickly as they progress through their shift: *inspection duration* decreases by 3.5% for each subsequent inspection conducted within a day (Column 1: *number of prior inspections today* $\beta = -0.035$, $p < 0.01$). For context, recall that our primary results indicate that each subsequent inspection during the day cites an average of 3.15% fewer violations. The model reported in Column 3 of Table C4 indicates that the net effect is that inspector citation pace decreases by 1.3% for each subsequent inspection of the day (*number of prior inspections today* $\beta = -0.013$, $p < 0.10$).

Turning to *potentially shift-prolonging* inspections, recall that our primary results indicated that these had 5.07% fewer citations. Column 1 of Table C4 reveals that inspectors conduct such inspections 3.6% more quickly (*potentially shift-prolonging* $\beta = -0.036$, $p < 0.01$). Column 3 reveals that the effect of *potentially shift-prolonging* on citation pace is not statistically significant. These results suggest that the diminishments in citations result from shorter inspection durations rather than slower inspector' speed, with inspectors' citation pace remaining largely unaffected by the risk of working beyond their typical end of shift. This suggests that our earlier finding that *potentially shift prolonging* inspections result in fewer violations is likely due to inspectors' desire to avoid working late, rather than fatigue eroding their citation pace.

Turning to potential outcome-effects, we find no evidence that the outcome of the inspector's prior inspection affects inspection duration, as the coefficients on the variables related to the prior inspected establishment's violations and violation trend are not statistically significant (Columns 1 and 2 of Table C4). Recall that our primary results found that more violations and worsening trends at an inspector's prior establishment predicted that the inspector cited more violations at the focal inspection. Results reported in Column 3 of Table C4 indicate that inspectors' citation pace increases by 1.0% for each additional violation at the prior establishment ($\beta = 0.010$, $p < 0.01$) and by 1.9% for each one-standard-deviation increase in the *prior inspected establishment's violation trend* ($\beta = 0.012$, $p < 0.05$). Column 4 indicates that, as was the case with the number of violations, the latter effect is asymmetric and driven by negative trends: whereas we find no change in inspectors' citation pace after inspecting an establishment that exhibited salient improvement, it does increase by 3.9% after inspecting an establishment that exhibited salient deterioration. This indicates that our earlier outcome-effect findings—that more violations and worsening trends at an inspector's prior establishment increase the inspector's citations at his or her next inspection—result from inspectors increasing their citation pace rather than spending more time onsite.

4.7.4. Violation Documentation. We investigate whether our hypothesized spillovers also affected the text comments that inspectors sometimes included to annotate violations. Fatigue might cause inspectors not to put as much effort into documenting the violations they cite, decreasing the average length of these comments. To assess documentation effort, we measure an inspection's *average violation comment length* as the number of characters[14] the inspector typed into the comment field accompanying each cited violation (empty comment fields had zero characters), averaged across all violations cited in

[14] Results are substantively similar if we instead use the number of words.

that inspection. This variable is coded as missing for inspections that had no cited violations. We also calculated versions of this metric separately for critical and noncritical violations, which we refer to as *average critical violation comment length* and *average noncritical violation comment length*. We estimate the log (after adding 1) of each of these measures on our primary models, restricting the sample to those inspections with at least one violation, one critical violation and one noncritical violation, respectively (Appendix Table C5).

The results yield no evidence of any type of daily-schedule effects on comment length, regardless of how we measured it. As for outcome-effects, we find no effect of *prior inspected establishment's violations* on any measure of comment length, but we do find that worsening violation trends at the inspector's prior establishment are associated with shorter comments except for noncritical violations. Specifically, a one-standard-deviation increase in the *prior inspected establishment's violation trend* is associated with an average 1.1% decrease in the number of violation comment words in the focal inspection (Column 7: $\beta = -0.007$, $p < 0.10$) and a comparable average 1.1% decrease in the comment length in terms of number of characters (Column 1: $\beta = -0.007$, $p < 0.10$). Decomposing the latter effect reveals an average 1.6% decrease in the comment length for critical violations (Column 3: $\beta = -0.010$, $p < 0.05$) and no average impact on the comment length for noncritical violations (Column 5). Columns 2 and 8 reveal that these effects are driven by decreases in comment length after inspectors observe salient deterioration, rather than increases in comment length after observing salient improvement.

Overall, considered along with our citation pace results, these documentation findings suggest that a potential mechanism by which outcome-effects increase citation pace (that is, improve inspectors' productivity in citing violations) is by focusing attention on detection at the (less costly) expense of reducing documentation. Because each violation citation references the regulatory code infringed and only on some occasions does customization of violation comments provide additional value, we interpret these results as suggesting that inspectors successfully redirect their attention to important matters.

5. Discussion

We find strong evidence that inspectors' scrutiny is affected by their daily schedules and their experience at the immediately preceding inspection. As inspectors conduct additional inspections throughout the day, their scrutiny is eroded by increasing fatigue and by the perceived time pressure to complete their final inspection of the day at their usual time. We also find strong evidence that inspector scrutiny is influenced by their experience at their prior inspected establishment.

5.1. Contributions

Our work contributes to three literature streams. First, this study is among the first to bring an operational lens to the literature on monitoring and assessment of standards adherence. In particular, we identify important scheduling effects on the scrutiny and thus the accuracy of those who monitor establishments' adherence to standards. Our analysis of the effects of the timing of an inspection during inspectors' daily schedules and of the spillover effects from the outcomes of prior inspections at other establishments complements this literature's focus on improving the effectiveness of monitoring schemes.

Second, by identifying spillover effects between inspections, our findings contribute to a related literature on the spillover effects of regulatory sanctions (e.g., Cohen 2000, Shimshack and Ward 2005). While that literature focuses on how an inspection agency's monitoring efforts and enforcement actions affect its reputation for stringency, which has a spillover influence on other establishments' compliance, our study focuses on how inspectors' experiences at one establishment have spillover effects on their scrutiny at subsequent inspections of other establishments. Ours is thus the first study of which we are aware that identifies spillover effects on inspector stringency associated not only with the outcomes of the immediately preceding inspection, but also with how many prior inspections an inspector had already conducted that day and with the inspector's apparent desire to avoid prolonging the workday. Moreover, our work contributes to the nascent literature on the accuracy of inspections—specifically, of regulatory regimes and third-party monitoring of labor conditions in supply chains—that has largely focused on how inspector bias is due to economic conflicts of interest or site-specific experience (e.g., Duflo et al. 2013, Short and Toffel 2016, Short, Toffel, and Hugill 2016, Ball, Siemsen, and Shah 2017). Our study is the first to our knowledge that brings the operational lens of scheduling to this literature by showing how work schedules can drive inaccuracies.

Third, our work also contributes to the literature on the performance implications of scheduling and task sequencing. By examining actual decisions with important consequences for public health, we contribute to the recent attempts to explore high-stakes decision making in field settings (e.g., Chen et al. 2016). While our findings may be consistent with status-quo bias increasing with the cumulative number of decisions over a day, we find that this bias manifests as inspectors becoming less stringent (by citing fewer violations), which contrasts with prior work that finds decision makers becoming more stringent over the course of the day (Danziger, Levav, and Avnaim-Pesso 2011). In our context, conducting additional food safety inspections throughout the day leads to mental and physical fatigue, eroding the inspector's ability to engage in physical effort and in social interactions with establishment staff (and the resulting impact on emotions and perceptions), which reduces violation detection and thus inspection quality. Our daily schedule effects findings complement the literature that has found that increased worker fatigue after long hours has led to accidents among nuclear and industrial plant operators, airline pilots, truck drivers, and hospital workers (Dinges 1995, Landrigan et al. 2004) and that workers seek to reduce their workload as they approach their end of shift (Chan 2015).[15] In response to such findings, industry standards and regulations have established caps on the number of consecutive work hours in some of these professions, but those rules remain controversial and their benefits are still being assessed. We contribute to this debate by providing evidence of the negative effects of fatigue on the quality of work during shorter periods (normal shifts, rather than the overwork examined in prior studies), in a different setting (health inspections), focusing on primary (rather than secondary) tasks, and by identifying potential remedies. Our results suggest that inspectors themselves seek to ameliorate the

[15] Chan (2015) finds that hospital emergency department physicians approaching their end of shift accept fewer patients and spend less time with them. Whereas that study focuses on fixed shifts where physicians can pass along unfinished work to the next shift, our study focuses on inspectors who work without fixed shifts, do not pass along unfinished work to others, and do not stop working on an unfinished inspection once it has begun.

effects of their schedules (for example, by focusing their attention on critical violations at the expense of detecting fewer noncritical violations and producing less documentation), yet less-demanding work hours would improve inspection accuracy.

In addition, our work responds to the call for behavioral research in the operations management field (Bendoly, Donohue, and Schultz 2006, Gino and Pisano 2008) by identifying ways in which task sequencing affects worker behavior. Our finding that inspectors' experiences at prior inspections bias their subsequent inspections shows that the outcome of tasks can affect how humans—unlike machines—perform their next task.

5.2. Managerial Implications

By revealing biases that result from inspector schedules, our findings highlight opportunities to mitigate bias through alternative scheduling policies. This has important implications for inspectors focusing on their own companies' processes, third-party inspectors hired by buyers to assess their suppliers, and government inspectors assessing compliance with regulations governing occupational safety, environmental conditions, food safety, and other topics. Managers can develop strategies to improve inspection accuracy by seeking to reduce both the extent to which these biases erode inspection accuracy and the consequences of these biases.

One way to reduce the extent to which these biases erode inspection accuracy is to impose a cap on the number of inspections conducted each day to limit the deleterious effects of fatigue, although this risks reducing inspection capacity. Another approach, which can be used at the same time, is to minimize the number of shift-prolonging inspections by reallocating an inspector's weekly schedule to reduce variation in the predicted completion time of their final inspection each day or by shifting administrative tasks (such as office meetings) from the beginning to the end of the day. Our computations suggests that, in our setting, eliminating these two factors would result in 4.5% more citations, which, translated into health benefits, correspond to 18% fewer hospitalizations for foodborne illness. Reorganization of inspectors' schedules could eliminate these negative outcomes.

Our identified outcome-effects imply that increasing the salience of noncompliance and the need to enforce regulation could increase the number of violations detected. This suggests that reminders or other ways to increase such salience could potentially be a lever for inspection managers to increase the stringency of inspectors, even if the information is already available to them and despite their innate desire to protect consumers.

Managers can also use our findings to develop policies to reduce the *consequences* of inspector biases eroding inspection accuracy. For example, understanding that scrutiny typically declines as inspectors (a) conduct successive inspections during the day and (b) conduct inspections that risk prolonging their shift, the inspectors themselves could be required to schedule establishments that pose greater risks earlier in the shift. By improving inspector effectiveness in the case of food safety, these changes could reduce risk to public health.

5.3. Limitations and Future Research

Our study has several limitations that could be explored in future research. Though our data contains details of inspections and citations, we do not observe inspectors' beliefs or their interactions with the establishment personnel. We find that inspectors cite fewer violations after inspecting establishments that had fewer violations. Perhaps they make less effort to find hidden violations and are more willing to take a coaching approach—emphasizing education over enforcement and training operators to operate with better hygiene for borderline violations—rather than writing citations. Possible extensions of our study could use observations of these actions to quantify their relative roles on the net effects. In addition, although our research context—food safety inspections—is common worldwide, it is just one of many types of inspection conducted by companies and governments. Future research should examine whether the relationships we identified hold in other contexts.

6. References

Anand G., Gray J., Siemsen E. 2011. Decay, shock, and renewal: Operational routines and process entropy in the pharmaceutical industry. *Organization Science* **23**(6) 1700-1716.

Azoulay P., Graff Zivin J.S., Wang J. 2010. Superstar extinction. *The Quarterly Journal of Economics* **125**(2) 549-589.

Ball G.P., Siemsen E., Shah R. 2017. *Inspector experience and product recalls in the medical device industry*. Working Paper.

Ballou D.P., Pazer H.L. 1982. The impact of inspector fallibility on the inspection policy in serial production systems. *Management Science* **28**(4) 387-399.

Barsade S.G. 2002. The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly* **47**(4) 644-675.

Batt R.J., Terwiesch C. 2016. Early task initiation and other load-adaptive mechanisms in the Emergency Department. *Management Science* (Forthcoming).

Baumeister R.F., Bratslavsky E., Finkenauer C., Vohs K.D. 2001. Bad is stronger than good. *Review of General Psychology* **5**(4) 323-370.

Bendoly E., Donohue K., Schultz K.L. 2006. Behavior in operations management: Assessing recent findings and revisiting old assumptions. *Journal of Operations Management* **24**(6) 737-752.

Bennett V.M., Pierce L., Snyder J.A., Toffel M.W. 2013. Customer-driven misconduct: How competition corrupts business practices. *Management Science* **59**(8) 1725-1742.

Berry Jaeker J.A., Tucker A.L. 2016. Past the point of speeding up: The negative effects of workload saturation on efficiency and patient severity. *Management Science*.

Brachet T., David G., Drechsler A.M. 2012. The effect of shift structure on performance. *American Economic Journal: Applied Economics* **4**(2) 219-246.

Cameron A.C., Trivedi P.K. 2010. *Microeconometrics using Stata, revised edition*. Stata Press.

Chan D.C., Jr. 2015. The efficiency of slacking off: Evidence from the emergency department. *National Bureau of Economic Research Working Paper Series* **No. 21002**.

Chen D.L., Moskowitz T.J., Shue K. 2016. Decision making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires. *The Quarterly Journal of Economics* **131**(3) 1181-1242.

Cohen M.A. 2000. Empirical research on the deterrent effect of environmental monitoring and enforcement. *Environmental Law Reporter News and Analysis* **30**(4) 10245-10252.

Corbett C.J. 2006. Global diffusion of ISO 9000 certification through supply chains. *Manufacturing & Service Operations Management* **8**(4) 330-350.

Dai H., Milkman K.L., Hofmann D.A., Staats B.R. 2015. The impact of time at work and time off from work on rule compliance: The case of hand hygiene in health care. *Journal of Applied Psychology* **100**(3) 846-862.

Danziger S., Levav J., Avnaim-Pesso L. 2011. Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences* **108**(17) 6889-6892.

Delasay M., Ingolfsson A., Kolfal B. 2016. Modeling load and overwork effects in queueing systems with adaptive service rates. *Operations Research* **64**(4) 867-885.

Deo S., Jain A. 2015. *Slow first, fast later: Empirical evidence of speed-up in service episodes of finite duration*. SSRN Working Paper.

Dinges D.F. 1995. An overview of sleepiness and accidents. *Journal of Sleep Research* **4** 4-14.

Duflo E., Greenstone M., Pande R., Ryan N. 2013. Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. *The Quarterly Journal of Economics* **128**(4) 1499-1545.

Feinstein J.S. 1989. The safety regulation of U.S. nuclear power plants: Violations, inspections, and abnormal occurrences. *Journal of Political Economy* **97**(1) 115-154.

Freeman M., Savva N., Scholtes S. 2017. Gatekeepers at work: An empirical analysis of a maternity unit. *Management Science*.

Gawande K., Bohara A.K. 2005. Agency problems in law enforcement: Theory and application to the U.S. Coast Guard. *Management Science* **51**(11) 1593-1609.

Gino F., Pisano G.P. 2008. Toward a theory of behavioral operations. *Manufacturing and Service Operations Management* **10**(4) 676-691.

Glaeser E.L., Hillis A., Kominers S.D., Luca M. 2016. Crowdsourcing city government: Using tournaments to improve inspection accuracy. *American Economic Review* **106**(5) 114-118.
Gourieroux C., Monfort A., Trognon A. 1984. Pseudo Maximum Likelihood methods: Applications to Poisson models. *Econometrica* **52**(3) 701-720.
Gray J.V., Anand G., Roth A.V. 2015. The Influence of ISO 9000 certification on process compliance. *Production and Operations Management* **24**(3) 369-382.
Gray J.V., Siemsen E., Vasudeva G. 2015. Colocation still matters: Conformance quality and the interdependence of R&D and manufacturing in the pharmaceutical industry. *Management Science* **61**(11) 2760-2781.
Green L.V., Savin S., Savva N. 2013. "Nursevendor problem": Personnel staffing in the presence of endogenous absenteeism. *Management Science* **59**(10) 2237-2256.
Handley S.M., Gray J.V. 2013. Inter-organizational quality management: The use of contractual incentives and Monitoring mechanisms with outsourced manufacturing. *Production and Operations Management* **22**(6) 1540-1556.
Harrington W. 1988. Enforcement leverage when penalties are restricted. *Journal of Public Economics* **37** 29–53.
Hasija S., Pinker E., Shumsky R.A. 2009. OM Practice—Work expands to fill the time available: Capacity estimation and staffing under Parkinson's Law. *Manufacturing & Service Operations Management* **12**(1) 1-18.
Helland E. 1998. The enforcement of pollution control laws: inspections, violations, and self-reporting. *Review of Economics and Statistics* **80**(1) 141–153.
Ibanez M.R., Clark J.R., Huckman R.S., Staats B.R. 2016. *Discretionary task ordering: Queue management in radiological services*. Harvard Business School Working Paper.
Jin G.Z., Leslie P. 2003. The effect of information on product quality: Evidence from restaurant hygiene grade cards. *The Quarterly Journal of Economics* **118**(2) 409-451.
Jin G.Z., Leslie P. 2009. Reputational incentives for restaurant hygiene. *American Economic Journal: Microeconomics* **1**(1) 237-267.
Johnson M.S., Levine D.I., Toffel M.W. 2017. *Can rules-based targeting improve regulatory effectiveness? Evidence applying machine learning to randomized safety inspections*.
Kahneman D. 2003. Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review* **93**(5) 1449--1475.
Kang J.S., Kuznetsova P., Choi Y., Luca M. 2013. *Where not to eat? Improving public policy by predicting hygiene inspections using online reviews*.
Kc D.S., Terwiesch C. 2009. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* **55**(9) 1486-1498.
Kc D.S., Terwiesch C. 2012. An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management* **14**(1) 50-65.
Kim S.-H. 2015. Time to come clean? Disclosure and inspection policies for green production. *Operations Research* **63**(1) 1-20.
Ko K., Mendeloff J., Gray W. 2010. The role of inspection sequence in compliance with the US Occupational Safety and Health Administration's (OSHA) standards: Interpretations and implications. *Regulation & Governance* **4**(1) 48-70.
Koh K., Rajgopal S., Srinivasan S. 2013. Non-audit services and financial reporting quality: evidence from 1978 to 1980. *Review of Accounting Studies* **18**(1) 1-33.
Kök A.G., Shang K.H. 2007. Inspection and replenishment policies for systems with inventory record inaccuracy. *Manufacturing & Service Operations Management* **9**(2) 185-205.
Kuntz L., Mennicken R., Scholtes S. 2015. Stress on the ward: Evidence of safety tipping points in hospitals. *Management Science* **61**(4) 754-771.
Lana F. 2003. Targeting enforcement to improve compliance with environmental regulations. *Journal of Environmental Economics and Management* **46** 72–85.

Landrigan C.P., Rothschild J.M., Cronin J.W., Kaushal R., Burdick E., Katz J.T., Lilly C.M., Stone P.H., Lockley S.W., Bates D.W., Czeisler C.A. 2004. Effect of reducing interns' work hours on serious medical errors in intensive care units. *New England Journal of Medicine* **351**(18) 1838-1848.

Langpap C., Shimshack J.P. 2010. Private citizen suits and public enforcement: substitutes or complements? *Journal of Environmental Economics and Management* **59** 235–249.

Lapré M.A., Mukherjee A.S., Van Wassenhove L.N. 2000. Behind the learning curve: Linking learning activities to waste reduction. *Management Science* **46**(5) 597-611.

Leape L.L. 2002. Reporting of adverse events. *New England Journal of Medicine* **347**(20) 1633-1638.

Lee H.L., Rosenblatt M.J. 1987. Simultaneous determination of production cycle and inspection schedules in a production systems. *Management Science* **33**(9) 1125-1136.

Lehman D.W., Kovács B., Carroll G.R. 2014. Conflicting social codes and organizations: Hygiene and authenticity in consumer evaluations of restaurants. *Management Science* **60**(10) 2602-2617.

Levine D.I., Toffel M.W. 2010. Quality management and job quality: How the ISO 9001 standard for quality management systems affects employees and employers. *Management Science* **56**(6) 978-996.

Levine D.I., Toffel M.W., Johnson M.S. 2012. Randomized government safety inspections reduce worker injuries with no detectable job loss. *Science* **336**(6083) 907-911.

Linder J.A., Doctor J.N., Friedberg M.W., et al. 2014. TIme of day and the decision to prescribe antibiotics. *JAMA Internal Medicine* **174**(12) 2029-2031.

Locke R.M. 2013. *The promise and limits of private power*. Cambridge University Press, Cambridge, UK.

Macher J.T., Mayo J.W., Nickerson J.A. 2011. Regulator heterogeneity and endogenous efforts to close the information asymmetry gap. *The Journal of Law & Economics* **54**(1) 25-54.

Marcora S.M., Staiano W., Manning V. 2009. Mental fatigue impairs physical performance in humans. *Journal of Applied Physiology* **106**(3) 857.

May P.J., Wood R.S. 2003. At the regulatory front lines: Inspectors' enforcement styles and regulatory compliance. *Journal of Public Administration Research and Theory* **13**(2) 117-139.

Minor T., Lasher A., Klontz K., Brown B., Nardinelli C., Zorn D. 2015. The per case and total annual costs of foodborne illness in the United States. *Risk Analysis* **35**(6) 1125-1139.

Muraven M., Baumeister R.F. 2000. Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin* **126**(2) 247-259.

Neuman J.H., Baron R.A. 1997. *Aggression in the workplace*. Sage Publications, Thousand Oaks, Calif.

Nickerson R.S. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* **2**(2) 175-220.

Oliva R., Sterman J.D. 2001. Cutting corners and working overtime: Quality erosion in the service industry. *Management Science* **47**(7) 894-914.

Pautz M.C. 2009. Trust between regulators and the regulated: A case study of environmental inspectors and facility personnel in Virginia. *Politics & Policy* **37**(5) 1047-1072.

Pautz M.C. 2010. Front-line regulators and their approach to environmental regulation in Southwest Ohio. *Review of Policy Research* **27**(6) 761-780.

Pierce L., Toffel M.W. 2013. The role of organizational scope and governance in strengthening private monitoring. *Organization Science* **24**(5) 1558-1584.

Porteus E.L., Angelus A. 1997. Opportunities for improved statistical process control. *Management Science* **43**(9) 1214-1228.

Powell A., Savin S., Savva N. 2012. Physician workload and hospital reimbursement: Overworked physicians generate less revenue per patient. *Manufacturing & Service Operations Management*.

Ronen J. 2010. Corporate audits and how to fix them. *Journal of Economic Perspectives* **24**(2) 189-210.

Rozin P., Royzman E.B. 2001. Negativity bias, negativity dominance, and contagion. *Personality & Social Psychology Review* **5**(4) 296-320.

Samuelson W., Zeckhauser R. 1988. Status quo bias in decision making. *Journal of Risk and Uncertainty* **1**(1) 7-59.

Scallan E., Griffin P.M., Angulo F.J., Tauxe R.V., Hoekstra R.M. 2011. Foodborne illness acquired in the United States—unspecified agents. *Emerging Infectious Disease journal* **17**(1) 16.
Scharff R.L. 2012. Economic burden from health losses due to foodborne illness in the United States. *Journal of Food Protection* **75**(1) 123-131.
Schultz K.L., Juran D.C., Boudreau J.W., McClain J.O., Thomas L.J. 1998. Modeling and worker motivation in JIT production systems. *Management Science* **44**(12-part-1) 1595-1607.
Shah R., Ball G.P., Netessine S. 2016. Plant operations and product recalls in the automotive industry: An empirical investigation. *Management Science*.
Shimshack J.P. 2014. The economics of environmental monitoring and enforcement. *Annual Review of Resource Economics* **6** 339–360.
Shimshack J.P., Ward M.B. 2005. Regulator reputation, enforcement, and environmental compliance. *Journal of Environmental Economics and Management* **50**(3) 519-540.
Short J.L., Toffel M.W. 2016. The integrity of private third-party compliance monitoring. *Administrative & Regulatory Law News* **42**(1) 22-25.
Short J.L., Toffel M.W., Hugill A.R. 2016. Monitoring global supply chains. *Strategic Management Journal* **37**(9) 1878–1897.
Simon P.A., Leslie P., Run G., Jin G.Z., Reporter R., Aguirre A., Fielding J.E. 2005. Impact of restaurant hygiene grade cards on foodborne-disease hospitalizations in Los Angeles County. *Journal of environmental health* **67**(7) 32-36, 56; quiz 59-60.
Simonsohn U., Gino F. 2013. Daily horizons: Evidence of narrow bracketing in judgment from 10 Years of M.B.A. admissions interviews. *Psychological Science* **24**(2) 219-224.
Staats B.R., Dai H., Hofmann D., Milkman K.L. 2016. Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare. *Management Science*.
Staats B.R., Gino F. 2012. Specialization and variety in repetitive tasks: Evidence from a Japanese bank. *Management Science* **58**(6) 1141-1159.
Stafford S.L. 2003. Assessing the effectiveness of state regulation and enforcement of hazardous waste. *Journal of Regulatory Economics* **23**(1) 27–41.
Tan T.F., Netessine S. 2014. When does the devil make work? An empirical study of the impact of workload on server's performance. *Management Science* **60**(6) 1574-1593.
Toffel M.W., Short J.L. 2011. Coming clean and cleaning up: Does voluntary self-reporting indicate effective self-policing? *The Journal of Law and Economics* **54**(3) 609-649.
Toffel M.W., Short J.L., Ouellet M. 2015. Codes in context: How states, markets, and civil society shape adherence to global labor standards. *Regulation & Governance* **9**(3) 205-223.
Tversky A., Kahneman D. 1974. Judgment under uncertainty: Heuristics and biases. *Science* **185**(4157) 1124-1131.
Wright R.A., Junious T.R., Neal C., Avello A., Graham C., Herrmann L., Junious S., Walton N. 2007. Mental fatigue influence on effort-related cardiovascular response: difficulty effects and extension across cognitive performance domains. *Motivation and Emotion* **31**(3) 219-231.

Table 1. Summary Statistics

Variable	Description	Mean	SD	Min	Max
Violations	Number of violations cited in the inspection	2.42	2.73	0	25
Number of prior inspections today	Number of inspections that the inspector had already conducted before the focal inspection on the same day	0.94	1.10	0	9
Potentially shift-prolonging	Indicates if the anticipated end time of an inspection (calculated as the inspection start time plus the duration of that establishment's previous inspection) falls after the inspector's running average daily clock-out time based on all of that inspector's preceding days in our sample	0.26	0.44	0	1
Prior inspected establishment's violations	Number of violations cited at the establishment inspected by the inspector immediately prior to the focal inspection	2.11	2.62	0	25
Prior inspected establishment's violation trend	Percentage change in the number of violations at that establishment between that day's inspection and its previous inspection (adding one to the denominator to avoid dividing by zero)	0.42	1.58	-0.95	23
Prior inspected establishment saliently improved	Indicates if the inspector's prior inspected establishment *improved saliently* (i.e., its current inspection yielded at least two *fewer* violations than its previous inspection)	0.24	0.43	0	1
Prior inspected establishment saliently deteriorated	Indicates if the inspector's prior inspected establishment *deteriorated saliently* (i.e, its current inspection yielded at least two *more* violations than its previous inspection)	0.21	0.41	0	1
Inspector experience	Number of inspections the inspector had conducted (at any establishment) since the beginning of our sample period by the time he or she began the focal inspection	520.09	303.30	1	1429
Returning inspector	Indicates if the inspector of the focal inspection had inspected the establishment beforehand	0.84	0.37	0	1
Establishment's nth inspection (second through tenth or more)	Indicators that indicate whether an inspection is the establishment's first, second, third (and so on) inspection in our sample period	4.04	2.15	1	20
Breakfast period (midnight to 10:59 am)	Indicates if the inspection began midnight to 10:59 am	0.32	0.47	0	1
Lunch period (11:00 am – 3:59 pm)	Indicates if the inspection began 11:00 am – 3:59 pm (omitted category)	0.66	0.47	0	1
Dinner period (4:00 pm–11:59 pm)	Indicates if the inspection began 4:00 pm–11:59 pm	0.02	0.15	0	1

N = 12,017 inspections

Table 2. Correlations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) Violations	1.00										
(2) Number of prior inspections today	-0.02	1.00									
(3) Potentially shift-prolonging	-0.01	0.36	1.00								
(4) Prior inspected establishment's violations	0.18	-0.03	0.03	1.00							
(5) Prior inspected establishment's violation trend	0.11	-0.01	0.01	0.56	1.00						
(6) Prior inspected establishment saliently improved	0.01	0.02	0.00	-0.17	-0.37	1.00					
(7) Prior inspected establishment saliently deteriorated	0.12	0.00	0.01	0.60	0.69	-0.29	1.00				
(8) Inspector experience	-0.05	0.05	-0.02	-0.01	0.01	-0.03	0.01	1.00			
(9) Returning inspector	-0.14	-0.01	0.10	-0.03	-0.01	-0.01	-0.01	0.33	1.00		
(10) Establishment's nth inspection (second through tenth or more)	0.05	0.00	0.06	0.01	0.02	-0.01	0.02	0.47	0.34	1.00	
(11) Breakfast period (midnight to 10:59 am)	0.01	-0.47	-0.41	-0.03	-0.02	0.01	-0.02	0.04	0.02	0.03	1.00
(12) Dinner period (4:00 pm–11:59 pm)	0.01	0.15	0.21	0.01	-0.01	0.01	0.00	-0.10	-0.07	-0.05	-0.10

N = 12,017 inspections

Table 3. How Inspectors' Schedules Influence Inspection Outcomes

	Dependent variable:	*violations*	
		(1)	(2)
H1	Number of prior inspections today	-0.032***	-0.032***
		(0.011)	(0.011)
H2	Potentially shift-prolonging	-0.052**	-0.051**
		(0.025)	(0.025)
H3	Prior inspected establishment's violations	0.015***	0.014***
		(0.004)	(0.004)
H4	Prior inspected establishment's violation trend	0.013**	
		(0.006)	
H5	Prior inspected establishment saliently improved		0.012
			(0.023)
H5	Prior inspected establishment saliently deteriorated		0.075***
			(0.027)
	Inspector experience	0.001***	0.001***
		(0.000)	(0.000)
	Returning inspector	-0.116***	-0.118***
		(0.035)	(0.035)
	Breakfast period (midnight to 10:59 am)	0.056**	0.056**
		(0.025)	(0.026)
	Dinner period (4:00 pm–11:59 pm)	0.000	-0.002
		(0.078)	(0.078)
	Month fixed effects	Included	Included
	Year fixed effects	Included	Included
	Establishment's nth inspection (second through tenth or more) fixed effects	Included	Included
	Inspection-type fixed effects	Included	Included
	Establishment x Inspector fixed effects	Included	Included
	Number of observations (inspections)	12,017	12,017

Notes: Poisson regression coefficients with robust standard errors clustered by establishment.
*** p < 0.01, ** p < 0.05, * p < 0.10.

Table 4. Estimates of Nationwide Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Impact on violations			Impact on health and associated costs				
	Average predicted number of violations cited per inspection	Percent change in average predicted number of violations cited per inspection compared to the status quo	Change in nationwide annual number of violations cited compared to the status quo	Percent change in foodborne illness hospitalizations compared to the status quo	Change in nationwide annual number of foodborne illness hospitalizations compared to the status quo	Change in nationwide annual number of foodborne illness cases compared to the status quo, in millions	Change in nationwide annual costs of foodborne illness cases compared to the status quo, in millions	
							Lower Estimate	Upper Estimate
Daily schedule effects								
1 Impact of eliminating *number of prior inspections today* effect	2.498	3.07%	74,370	-12.27%	-15,711	-5.87	-$4,383	-$9,540
2 Impact of eliminating *potentially shift-prolonging* effect	2.457	1.36%	33,020	-5.45%	-6,976	-2.60	-$1,946	-$4,236
3 Impact of eliminating both *number of prior inspections today* and *potentially shift-prolonging* effects	2.533	4.50%	109,140	-18.01%	-23,056	-8.61	-$6,432	-$14,000
Outcome effects								
4 Impact of eliminating *prior inspected establishment's violations* effect	2.330	-3.88%	-94,110	15.53%	19,881	7.42	$5,546	$12,072
5 Impact of eliminating *prior inspected establishment's violation trend* effect	2.405	-0.77%	-18,750	3.09%	3,961	1.48	$1,105	$2,405
6 Impact of eliminating both *prior inspected establishment's violations* and *prior inspected establishment's violation trend* effects	2.313	-4.57%	-110,800	18.29%	23,407	8.74	$6,529	$14,213

Column 1 is the average predicted number of violations per inspection, based on Model 1 of Table 4, under each scenario.

Column 2 is calculated as the percent change in the average predicted number of violations per inspection, comparing each scenario (Column 1) to the status quo value of 2.42365.

Column 3 is calculated as the difference in the average predicted number of violations per inspection, comparing each scenario (Column 1) to the status quo value of 2.42365, and multiplying this by the one million inspections that are conducted nationwide each year.

Column 4 is calculated by multiplying the percent change in average predicted number of violations compared to the status quo (Column 2) by the ratio of the change in hospitalizations to the change in compliance (derived from the 20% hospitalizations decline per 5% improvement in restaurant compliance relationship reported by Jin and Leslie (2003), that is, -20%/5% = -4).

Column 5 is calculated as the difference in hospitalizations between (a) the estimated number that would have occurred had the effects in this row not influenced hospitalizations and (b) the 128,000 that actually occurred. Specifically, we multiply the percent change in hospitalizations (Column 4) by the 128,000 nationwide annual hospitalizations.

Column 6 is calculated by multiplying the change in nationwide annual number of foodborne illness hospitalizations compared to the status quo (Column 5) by the number of illness cases per hospitalization (derived from the ratio of the estimated 47.8 million annual foodborne illnesses to the resulting 128,000 hospitalizations by Scallan (2011), that is, 47.8 million/128,000).

Columns 7 and 8 are calculated by multiplying the estimated change in illness cases (Column 6) by $747 (the weighted average from Minor et al (2015)) and $1,626 (the enhanced model estimate from Scharff (2012)) in estimated costs per illness case, respectively.

Appendix A. Sample Construction and Restrictions

Part 1. Sample Construction

This appendix specifies how we converted the raw data into the estimation sample of 12,017 inspections of 3,399 establishments conducted on 6,880 inspector-days. The process consists of dropping duplicate records, correcting a.m./p.m. coding errors, constructing our measures, flagging remaining inspections with data errors and dropping those inspector-days, and omitting inspections that are dropped by our specification.

Step 1. Dropping duplicate records. We began by correcting several types of problems in the raw data. We first removed duplicate inspection records that corresponded to the same establishment and starting and ending times. For each of these pairs or triplets, we kept the observation that reported the most violations because our interviews with inspectors indicated that the others likely represented unfinished reports accidently submitted prior to the inspection being completed. This excluded 46 inspection records and resulted in 38,065 inspections included in the raw data that Hazel Analytics provided.

Because our specification results in an estimation sample that omits (a) inspector-establishment dyads that contain a single inspection, (b) inspector-establishment dyads with multiple inspections but all of which cited no violations, and (c) inspections for which the *prior inspected establishment's violation trend* (and thus, *prior inspected establishment saliently improved* and *prior inspected establishment saliently deteriorated*) is not available, our potential estimation sample is limited to 14,965 inspections. We use all inspections to construct variables using the full histories of inspectors and establishments; for example, we compute the variables related to the inspector's prior inspection using the inspector's true prior inspection even if it took place at an establishment inspected only once by this inspector and was therefore excluded from the estimation sample. We take the following steps to address data errors and further refine the sample.

Step 2. Correcting a.m./p.m. coding errors. We then attempted to correct instances in which inspectors miscoded an inspection's ending time as preceding its starting time. Our interviews with inspectors indicated that these instances very likely resulted from inspectors miscoding the a.m./p.m. field, which we corrected for 412 inspections (of which only 58 remain in our estimation sample) based on the following four scenarios:

(1) When the ending time was coded as preceding its starting time, the start time was coded as p.m. and the ending time was coded as a.m., and the hour (0-11) of the starting time was *higher* than the hour of the ending time, we changed the start and ending times to a.m. and p.m., respectively. For example, if an inspection's starting and ending times in the raw data appeared as 11 p.m. to 3 a.m., we corrected them to be 11 a.m. to 3 p.m.

(2) When the ending time was coded as preceding its starting time, the start time was coded as p.m. and the ending time was coded as a.m., and the hour (0-11) of the starting time was *lower* than the hour of the ending time, we manually examined each record and changed either the start time to a.m. or the ending time to p.m., depending on which of the two changes seemed more reasonable in the context of the inspector's day and full inspection history.

(3) When the ending time was coded as preceding its starting time and both start and ending times were *coded as a.m.*, we changed the ending time to p.m. For example, if an inspection's starting and ending times in the raw data appeared as 10 a.m. to 2 a.m., we corrected the ending time to 2 p.m.

(4) When the ending time was coded as preceding its starting time and both start and ending times were *coded as p.m.*, we changed the start time to a.m. For example, if the starting and ending times were coded as 10 p.m. and 1 p.m., we changed the starting time to 10 a.m.

We then examined another type of data entry error based on potential a.m./p.m. coding mistakes that led to start and/or end times occurring extremely early or late in the day—or in that shift—given that inspector's routine starting and ending times. Considering the context of the inspector-day and the inspector's full inspection history, we corrected the a.m./p.m. field for starting or ending times for 63 inspections (10 in our estimation sample).

Finally, we reviewed the inspections that started in the a.m., ended in the p.m., and had a duration of more than six hours (e.g., starting at 7 a.m. and ending at 8 p.m.), as this indicated possible data errors. Using the approach described earlier, we corrected the a.m./p.m. field for starting or ending times for 163 inspections (17 in our estimation sample).

Altogether, these steps led to corrected a.m./p.m. values for 638 inspections from the raw data of 38,065 inspections, which corresponds to only 85 inspections in our final estimation sample of 12,017 inspections.

Step 3. Construct measures. We then used the dataset with all the inspections to construct the measures, including those involving inspector sequences and lags.

Step 4. Flag remaining inspections with data errors and drop those inspector-days. We then identified inspections for which we could not adequately calculate relevant variables and dropped all inspector-days that contained any of these inspections because those inspector-days would not reflect the complete daily account of an inspector's inspections. Specifically, we flagged inspections whose times overlapped with other inspections conducted by that inspector. Examples include a seafood market and its sushi bar, a coffee place and deli service located in the same grocery store, or different establishments in a food court. We then flagged additional inspections whose end time was coded as preceding or equaling the start time. (This does not include those with a starting- or ending-time problem that was corrected, as described earlier, by adjusting the a.m./p.m. field.) We also flagged inspections that were coded as lasting more than 4.5 hours, which corresponds to the 99.4 percentile of inspection duration. To be conservative, we sought to estimate our model on only those inspector-days for which we had a complete daily account of the inspector's inspections. Accordingly, we flagged not only the 4,726 inspections affected, but also the 2,350 additional inspections that were conducted on the inspector-days during which at least one flagged inspection took place. In other words, if one inspection had to be dropped, the whole inspector-day was dropped. Accordingly, we excluded the 2,320 inspections that were flagged in our potential estimation sample of 14,965 inspections.

Step 5. Omit inspections dropped by our specification. As described earlier, our specification drops inspector-establishment dyads that either (a) contain a single observation, or (b) yielded zero violations in all inspections. Because of the changes made in the prior step, these restrictions led to further exclusion of 628 inspections, resulting in a final estimation sample of 12,017 inspections of 3,399 establishments conducted on 6,880 inspector-days.

Part 2. Results without Sample Restrictions

These sample restrictions do not affect our inferences, as all of our hypothesized results continue to hold when using alternative specifications that use the full sample of inspections, that is, all observations in the raw dataset except the 46 duplicates believed to be data errors. Because our main Poisson specification automatically restricts the sample, we also estimated the primary models with ordinary least squares regression predicting the log of violations (after adding 1) using the full sample. Table A1 reports the results, which confirm the robustness of our findings. Table A2 provides the summary statistics for this full sample.

Table A1. Robustness Tests: Ordinary Least Squares Estimation Using Full Sample

	Dependent variable:	log (*violations* + 1)			
		(1)	(2)	(3)	(4)
H1	Number of prior inspections today	-0.025***	-0.023***	-0.023***	-0.023***
		(0.004)	(0.004)	(0.005)	(0.005)
H2	Potentially shift-prolonging	-0.046***	-0.049***	-0.044***	-0.044***
		(0.011)	(0.011)	(0.014)	(0.013)
H3	Prior inspected establishment's violations		0.012***	0.010***	0.009***
			(0.002)	(0.003)	(0.003)
H4	Prior inspected establishment's violation trend			0.010**	
				(0.004)	
H5	Prior inspected establishment saliently improved				-0.005
					(0.012)
H5	Prior inspected establishment saliently deteriorated				0.039**
					(0.016)
	Inspector experience	0.000***	0.000***	0.000***	0.000***
		(0.000)	(0.000)	(0.000)	(0.000)
	Returning inspector	-0.131***	-0.130***	-0.102***	-0.102***
		(0.016)	(0.016)	(0.018)	(0.018)
	Breakfast period (midnight to 10:59 am)	0.008	0.008	0.016	0.016
		(0.011)	(0.011)	(0.013)	(0.013)
	Dinner period (4:00 pm–11:59 pm)	0.011	0.010	0.019	0.019
		(0.026)	(0.026)	(0.036)	(0.036)
	Month fixed effects	Included	Included	Included	Included
	Year fixed effects	Included	Included	Included	Included
	Establishment's nth inspection (second through tenth or more) fixed effects	Included	Included	Included	Included
	Inspection-type fixed effects	Included	Included	Included	Included
	Establishment x Inspector fixed effects	Included	Included	Included	Included
	Number of observations (inspections)	38,065	37,941	26,190	26,190
	R-squared	0.26	0.26	0.29	0.29

Notes: Ordinary least squares regression coefficients with robust standard errors clustered by establishment. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A2. Summary Statistics for Full Sample (38,065 inspections)

		Obs	Mean	Std. dev.	Min	Max
(1)	Violations	38,065	1.95	2.43	0	25
(2)	Number of prior inspections today	38,065	1.03	1.34	0	17
(3)	Potentially shift-prolonging	38,065	0.21	0.41	0	1
(4)	Prior inspected establishment's violations	37,941	1.95	2.43	0	25
(5)	Prior inspected establishment's violation trend	26,190	0.36	1.44	-0.95	23
(6)	Prior inspected establishment saliently improved	26,190	0.23	0.42	0	1
(7)	Prior inspected establishment saliently deteriorated	26,190	0.20	0.40	0	1
(8)	Inspector experience	38,065	359.01	318.45	0	1437
(9)	Returning inspector	38,065	0.47	0.50	0	1
(10)	Establishment's inspection sequence	38,065	2.83	1.96	1	20
(11)	Breakfast period (midnight to 10:59 am)	38,065	0.33	0.47	0	1
(12)	Dinner period (4:00 pm–11:59 pm)	38,065	0.05	0.23	0	1

Appendix B

Table B1. Supplemental Descriptive Statistics

Panel A. Inspection sequence within the day

1st inspection of the day	5,328
2nd inspection of the day	3,618
3rd inspection of the day	1,971
4th inspection of the day	763
5th inspection of the day	248
6th inspection of the day	61
7th+ inspection of the day	28
Total number of inspections:	12,017

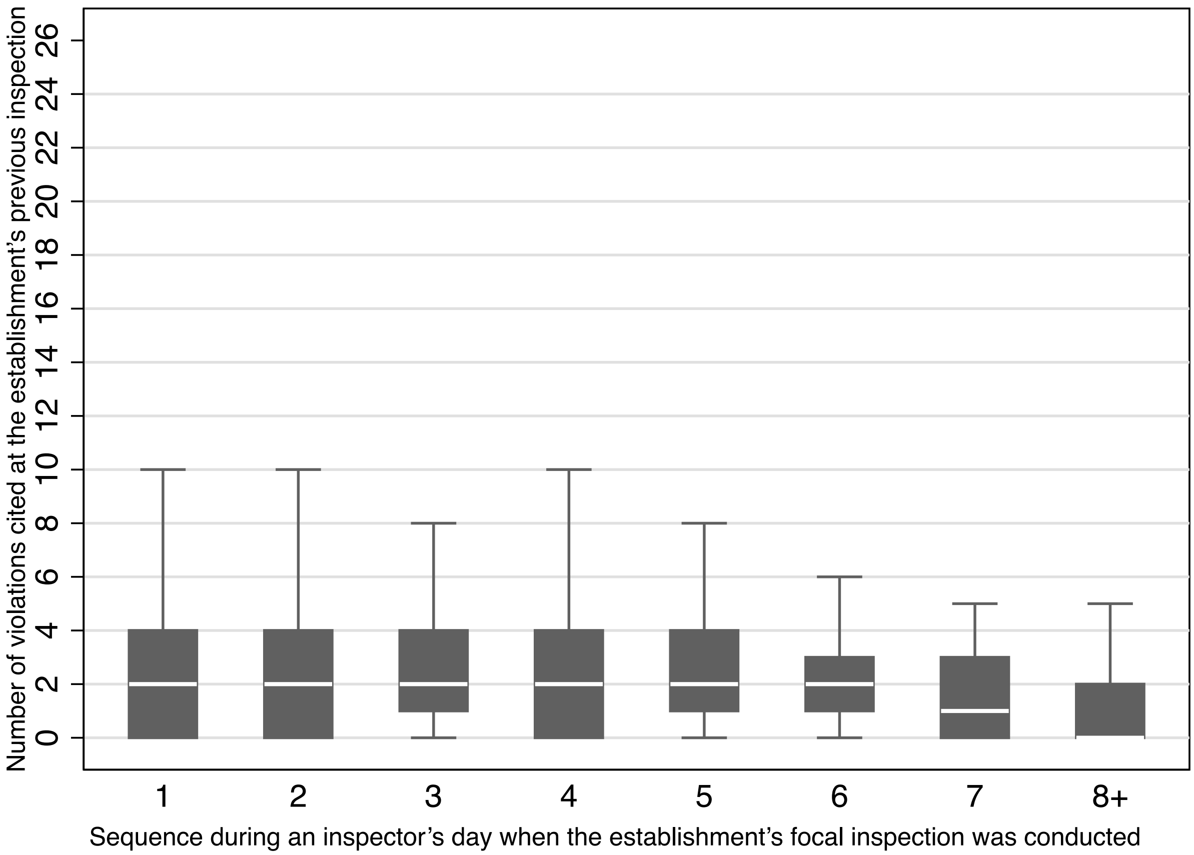
Panel B. Number of inspector-days

1 inspection days	1,790
2 inspection days	2,226
3 inspection days	1,637
4 inspection days	801
5 inspection days	295
6 inspection days	83
7+ inspection days	48
Total number of inspector-days:	6,880

Panel C. Inspections by hour begun and corresponding meal period

7 am or earlier	39	3,856 during breakfast period
8 am	222	
9 am	972	
10 am	2,623	
11 am	1,986	7,888 during lunch period
12 pm	1,331	
1 pm	2,331	
2 pm	1,653	
3 pm	587	
4 pm	171	273 during dinner period
5 pm	59	
6 pm or later	43	
Total number of inspections:	12,017	

Figure B1. The stability in medians and interquartile-ranges of establishments' previous inspection outcomes across sequential inspections within a day suggests that inspectors did not create their daily schedule based on establishments' previous violation counts.



Note. This figure displays a box-and-whisker plot of the number of violations cited at each establishment's *previous* inspection, categorized by when during an inspector's daily schedule its *subsequent* (that is, focal) inspection took place. The box represents the interquartile range segmented by the median. The adjacent lines extend above and below the box by 1.5 times the value of the interquartile range.

Appendix C

Table C1. Extensions: Summary Statistics

		Obs	Mean	Std. dev.	Min	Max
(1)	Penultimate inspected establishment's violations	12,014	2.10	2.60	0	25
(2)	Antepenultimate inspected establishment's violations	12,008	2.07	2.57	0	25
(3)	Prior inspected establishment's violations for the first inspection of the day	12,017	0.96	2.05	0	25
(4)	Prior inspected establishment's violations for the second+ inspection of the day	12,017	1.15	2.21	0	23
(5)	Critical violations	12,017	0.93	1.27	0	11
(6)	Noncritical violations	12,017	1.49	1.94	0	16
(7)	log inspection duration	12,017	4.01	0.55	1.61	5.60
(8)	log violation citations per hour	12,017	0.96	0.71	0	3.76
(9)	log average violation comment length	8,938	4.66	0.63	0	7.90
(10)	log average critical violation comment length	6,083	4.89	0.64	0	7.16
(11)	log average noncritical violation comment length	7,119	4.43	0.59	0	7.90
(12)	log average number of words in violation comments	8,938	2.94	0.58	0	6.06
(13)	Inspection duration (minutes)	12,017	62.87	32.01	5	270
(14)	Violation citations per hour	12,017	2.35	2.66	0	42
(15)	Average violation comment length	8,938	130.83	112.68	0	2687
(16)	Average critical violation comment length	6,083	164.28	125.42	0	1,287.50
(17)	Average noncritical violation comment length	7,119	102.28	93.06	0	2687
(18)	Average number of words in violation comments	8,938	21.96	18.35	0	429

Table C2. Persistence of Outcome-effects

Dependent variable:	*violations*		
	(1)	(2)	(3)
Number of prior inspections today	-0.030***	-0.029***	-0.031***
	(0.011)	(0.011)	(0.012)
Potentially shift-prolonging	-0.057**	-0.058**	-0.052**
	(0.025)	(0.025)	(0.025)
Prior inspected establishment’s violations	0.015***	0.015***	
	(0.004)	(0.004)	
Penultimate inspected establishment’s violations	0.010***	0.009***	
	(0.003)	(0.003)	
Antepenultimate inspected establishment’s violations		0.006*	
		(0.004)	
Prior inspected establishment’s violations			0.016***
for the first inspection of the day			(0.006)
Prior inspected establishment’s violations			0.015***
for the second+ inspection of the day			(0.005)
Prior inspected establishment’s violation trend	0.013**	0.013**	0.013**
	(0.006)	(0.006)	(0.006)
Inspector experience	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)
Returning inspector	-0.119***	-0.121***	-0.116***
	(0.034)	(0.034)	(0.035)
Breakfast period (midnight to 10:59 am)	0.055**	0.056**	0.055**
	(0.026)	(0.026)	(0.026)
Dinner period (4:00 pm–11:59 pm)	-0.004	-0.008	-0.000
	(0.077)	(0.077)	(0.078)
Month fixed effects	Included	Included	Included
Year fixed effects	Included	Included	Included
Establishment's nth inspection (second through tenth or more) fixed effects	Included	Included	Included
Inspection-type fixed effects	Included	Included	Included
Establishment x Inspector fixed effects	Included	Included	Included
Number of observations (inspections)	12,011	12,000	12,017

Notes: Poisson regression coefficients with robust standard errors clustered by establishment. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C3. Critical and Noncritical Violations

	Dependent variable:	*critical violations*		*noncritical violations*	
		(1)	(2)	(3)	(4)
H1	Number of prior inspections today	-0.014	-0.014	-0.042***	-0.041***
		(0.015)	(0.015)	(0.013)	(0.013)
H2	Potentially shift-prolonging	-0.037	-0.036	-0.060**	-0.059**
		(0.036)	(0.036)	(0.030)	(0.029)
H3	Prior inspected establishment's violations	0.018***	0.015***	0.014***	0.013***
		(0.006)	(0.006)	(0.005)	(0.005)
H4	Prior inspected establishment's violation trend	0.013*		0.012*	
		(0.008)		(0.007)	
H5	After salient improvement		-0.016		0.029
			(0.031)		(0.028)
H5	After salient deterioration		0.071*		0.074**
			(0.039)		(0.031)
	Breakfast period	0.046	0.046	0.063**	0.063**
	(midnight to 10:59 am)	(0.035)	(0.035)	(0.030)	(0.030)
	Dinner period	0.041	0.040	-0.039	-0.041
	(4:00 pm–11:59 pm)	(0.097)	(0.097)	(0.100)	(0.100)
	Inspector experience	0.001***	0.001***	0.001***	0.001***
		(0.000)	(0.000)	(0.000)	(0.000)
	Previous inspector	-0.113**	-0.114**	-0.105**	-0.107***
		(0.049)	(0.049)	(0.041)	(0.041)
	Month fixed effects	Included	Included	Included	Included
	Year fixed effects	Included	Included	Included	Included
	Establishment's nth inspection (second through tenth or more) fixed effects	Included	Included	Included	Included
	Inspection-type fixed effects	Included	Included	Included	Included
	Establishment x Inspector fixed effects	Included	Included	Included	Included
	Number of observations (inspections)	10,298	10,298	10,624	10,624

Notes: Poisson regression coefficients with robust standard errors clustered by establishment.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C4. Effects of Inspectors' Schedules on Speed and Citation Pace

Dependent variable:	Inspector speed log *inspection duration*		Inspector citation pace log (*violation citations per hour* + 1)	
	(1)	(2)	(3)	(4)
Number of prior inspections today	-0.035***	-0.035***	-0.013*	-0.014*
	(0.005)	(0.005)	(0.008)	(0.008)
Potentially shift-prolonging	-0.036***	-0.036***	-0.025	-0.024
	(0.011)	(0.011)	(0.018)	(0.018)
Prior inspected establishment's violations	0.002	0.002	0.010***	0.009***
	(0.002)	(0.002)	(0.003)	(0.003)
Prior inspected establishment's violation trend	-0.001		0.012**	
	(0.003)		(0.005)	
Prior inspected establishment saliently improved		0.010		-0.017
		(0.009)		(0.016)
Prior inspected establishment saliently deteriorated		0.004		0.038*
		(0.012)		(0.020)
Inspector experience	0.000	0.000	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Returning inspector	0.079***	0.079***	-0.100***	-0.100***
	(0.016)	(0.016)	(0.024)	(0.024)
Breakfast period (midnight to 10:59 am)	0.036***	0.036***	-0.013	-0.013
	(0.011)	(0.011)	(0.018)	(0.018)
Dinner period (4:00 pm–11:59 pm)	-0.036	-0.036	0.053	0.051
	(0.032)	(0.032)	(0.053)	(0.053)
Month fixed effects	Included	Included	Included	Included
Year fixed effects	Included	Included	Included	Included
Establishment's nth inspection (second through tenth or more) fixed effects	Included	Included	Included	Included
Inspection-type fixed effects	Included	Included	Included	Included
Establishment x Inspector fixed effects	Included	Included	Included	Included
Number of observations (inspections)	12,017	12,017	12,017	12,017
R-squared	0.45	0.45	0.20	0.20

Notes: Ordinary least squares coefficients with robust standard errors clustered by establishment.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table C5. Effects of Inspectors' Schedules on Documentation

Dependent variable:	log (*average violation comment length* + 1)		log (*average critical violation comment length* + 1)		log (*average noncritical violation comment length* + 1)		log (*average number of words in violation comments* + 1)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of prior inspections today	-0.002	-0.002	0.001	0.001	-0.002	-0.002	-0.002	-0.002
	(0.007)	(0.007)	(0.010)	(0.010)	(0.007)	(0.007)	(0.006)	(0.006)
Potentially shift-prolonging	0.003	0.003	0.022	0.022	-0.015	-0.015	0.008	0.008
	(0.015)	(0.015)	(0.024)	(0.024)	(0.016)	(0.016)	(0.014)	(0.014)
Prior inspected establishment's violations	0.001	0.001	0.003	0.003	-0.000	-0.000	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)
Prior inspected establishment's violation trend	-0.007*		-0.010**		0.000		-0.007*	
	(0.004)		(0.005)		(0.004)		(0.004)	
Prior inspected establishment saliently improved		-0.013		-0.011		-0.006		-0.009
		(0.014)		(0.021)		(0.016)		(0.013)
Prior inspected establishment saliently deteriorated		-0.032*		-0.038		-0.002		-0.033**
		(0.016)		(0.025)		(0.016)		(0.015)
Inspector experience	0.000	0.000	0.000	0.000	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Returning inspector	-0.020	-0.019	-0.010	-0.008	-0.030	-0.030	-0.017	-0.017
	(0.020)	(0.020)	(0.030)	(0.030)	(0.023)	(0.023)	(0.018)	(0.018)
Breakfast period (midnight to 10:59 am)	-0.024	-0.025	-0.007	-0.006	-0.029	-0.029	-0.020	-0.020
	(0.016)	(0.016)	(0.023)	(0.023)	(0.018)	(0.018)	(0.015)	(0.015)
Dinner period (4:00 pm–11:59 pm)	0.025	0.026	-0.052	-0.051	0.034	0.034	0.006	0.006
	(0.048)	(0.048)	(0.064)	(0.064)	(0.057)	(0.057)	(0.042)	(0.042)
Month fixed effects	Included	Included	Included	Included	Included	Included	Included	Included
Year fixed effects	Included	Included	Included	Included	Included	Included	Included	Included
Establishment's nth inspection (second through tenth or more) fixed effects	Included	Included	Included	Included	Included	Included	Included	Included
Inspection-type fixed effects	Included	Included	Included	Included	Included	Included	Included	Included
Establishment x Inspector fixed effects	Included	Included	Included	Included	Included	Included	Included	Included
Observations (inspections)	8,938	8,938	6,083	6,083	7,119	7,119	8,938	8,938
R-squared	0.04	0.04	0.05	0.05	0.05	0.05	0.04	0.04

Notes: Ordinary least squares coefficients with robust standard errors clustered by establishment.
*** p < 0.01, ** p < 0.05, * p < 0.10.